ESSAYS ON THE ECONOMICS OF INTERNET ENABLED MARKETS: LAST MILE INTERNET, E-COMMERCE, AND DIGITAL DIVIDE

A Dissertation
Presented to
The Academic Faculty

by

Karthik Babu Nattamai Kannan

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Scheller College of Business

Georgia Institute of Technology
August 2018

ESSAYS ON THE ECONOMICS OF INTERNET ENABLED MARKETS: LAST MILE INTERNET, E-COMMERCE, AND DIGITAL DIVIDE

Approved by:

Dr. Sridhar Narasimhan, Co-advisor
Scheller College of Business
Georgia Institute of Technology

Dr. Yu Jeffrey Hu, Co-advisor
Scheller College of Business
Georgia Institute of Technology

Dr. Eric Overby
Scheller College of Business
Georgia Institute of Technology

Dr. Sabyasachi Mitra
Scheller College of Business
Georgia Institute of Technology

Dr. Chris Forman
Dyson School of Applied Economics and Management
Cornell University

Date Approved: June 08, 2018

To my dear wife Sumana and lovely boys Srijan and Putush.

ACKNOWLEDGEMENTS

I would like to express my gratitude to members of my dissertation committee for their patience, support and exemplary guidance. I am foremost grateful to Dr. Sridhar Narasimhan for being an awesome mentor. A humble and positive force throughout this journey, Sri continues to inspire me with his curiosity, eagerness to learn and dynamic personality. I am extremely grateful to Dr. Jeffrey Hu for laying a strong foundation to execute research, for raising the bar, and for giving me plenty of opportunities to succeed in both research and teaching. I am forever grateful to Dr. Eric Overby for being open to my research ideas and shaping my job market paper to bring out the best in me. I thank Dr. Saby Mitra for his thought provoking and fundamental questions that has shaped my outlook towards research. My sincere thanks to Dr. Chris Forman for teaching me the foundations of economics of digitization, for being so generous to young scholars like me.

Many special thanks to all the faculty members, the fellow Ph.D. students, and alumni in the Information Technology Management department. Many thanks to Ursula Reynolds and our ITM PhD Coordinator Dr. Marius Florin Niculescu for building a strong student-centered PhD program at GT.

I thank the anonymous telecom company that supported two major parts of my dissertation and other research projects. I am forever indebted to my wife Sumana Sen for seeing the potential in me and encouraging me to pursue this PhD. I could not have completed this journey without her unconditional love and numerous sacrifices.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Improvements in technology such as internet, mobile phones, electronic markets have helped us make enormous social and economic progress. However, there is evidence of growing inequality between those who can use these technologies to their advantage and those who don't. In this dissertation, I examine how the internet enabled markets can potentially address three such problems. First, geographic divide – the uneven development between those who are in urban areas vs. those who are in the rural areas. Second, socioeconomic divide – the uneven development between those who have high education, income levels and those who don't. Third, the ability for (small scale) designers to compete against well-known brands and retail giants like Amazon.

In the first chapter, we examine whether improving mobile internet access is an effective method for closing the digital divide caused by geographic location or socioeconomic status. We do this by studying the adoption and use of unlimited mobile data plans offered by a large telecommunications provider. We find that adoption of an unlimited plan leads to a substantial increase in a household's data consumption, with the increase being particularly large for rural households and those of low socioeconomic status. This suggests that unlimited plans help these households "catch up", potentially narrowing the digital divide. Although most of the increase is accounted for by media and entertainment content, there is a significant increase in consumption of content likely to be socially beneficial: specifically news, education, and career-related content. We conclude that policy makers should encourage unlimited mobile data plans as a method to close the digital divide. They should also invest in educational programs on how to use the internet

for beneficial purposes and help web site providers make their services accessible via mobile phones.

The emergence of the maker movement and e-commerce platforms such as Etsy, Fab has democratized the production of goods and increased consumer demand for designer goods. However, designers with little brand recognition must overcome high buyer search costs to succeed in these markets or on their own websites. The second chapter examines two mechanisms that designers use to promote their products—flash sales and social media. Using two different identification techniques, we find that social media activities such as Facebook Likes, Pinterest Pins, and Faves have positive but different magnitudes of causal (and predictive) impact on sales, moderated by product type. Also, flash sale promotions increase average daily sales on the designer's primary website by up to 0.8 units in the first week after a flash sale is initiated.

The third chapter examines how promotion of free Wi-Fi hotspots impacts both paid mobile and free Wi-Fi data usage. Interestingly, we find promoting Wi-Fi hotspots leads to a small but significant mobile data usage, with the heavy mobile users having the highest impact. Also, increase in Wi-Fi usage is moderated by the type of business (or location) that provides the hotspots. Public places like airports, convention centers where people spend a lot of time have the highest increase in Wi-fi usage. Overall, our study reveals how and where Wi-Fi last mile channel can complement mobile internet usage.

We find that internet is a tool that has the potential to reduce the barrier between the haves and have nots. However, policy makers, managers and individuals have to understand the economics, mechanisms, and limitations of this tool in order to effectively utilize these technologies.

CHAPTER 1. INTRODUCTION

> "When the bases of the acquisition and distribution of goods are relatively stable, stratification by status is favored. Every technological repercussion and economic transformation threatens stratification by status and pushes the class situation into the foreground."
>
> Weber (2013)

Technological breakthroughs such as mobile phone, internet etc. have the potential to disrupt established social and economic structures by providing level playing field for both newcomers and existing players. For example, by reducing buyer search costs electronic markets (Bakos 1997, Malone et al. 1987) might help small scale designers compete with established brands and players like Amazon. By providing open access to hundreds of courses to anyone who has internet, MOOCS promise bridge the inequality caused by lack of education. The central question I ask in my dissertation is – *does the internet really achieve its promise of bridging social and economic inequality*?

While we have ample evidence on the positive effects of internet (Atasoy 2013, Greenstein and McDevitt 2011, Nattamai Kannan et al. 2016), many scholars (Hargittai 2008, Tichenor et al. 1970) have found that new technology breakthroughs merely reinforce pre-existing social and economic hierarchy. DiMaggio et al. (2004) find increasing digital inequality despite vast improvements in telecommunication technologies. Bonfadelli (2002) finds widening knowledge gap despite increased broadband usage. Forman et al. (2005) and Zook (2000, 2001) find increasing geographic concentration of firms in large urban areas despite the promise of internet to break locational barriers.

For example, Massively Online Open Courses (MOOCS) exploded in popularity in recent years with the promise of opening high quality courses taught by the most elite professors to the entire world free of charge. However, there is strong evidence that (Christensen et al. 2013) such courses are serving the well-educated, employed individuals mostly from developed countries instead of making a difference in the lives of those without access to higher education. Prior (2005) finds that despite abundant supply of political information through TV and the Internet, there is no significant difference in the political knowledge and turnout in elections. Even with greater choice and availability of free useful information, people tend to choose content that mirrors their *pre-existing* preferences.

In this dissertation, I examine how the internet can potentially address three such inequalities. First, geographic divide – the uneven development between those who are in urban areas vs. those who are in the rural areas. Second, socioeconomic divide – the uneven development between those who have high education, income levels and those who don't. the ability for (small scale) designers to compete against well-known brands and retail giants like Amazon.

Access to the internet has become a de facto requirement for participating in our increasingly digital society. Unfortunately, many households still struggle with limited internet access, particularly those in rural areas and of low socioeconomic status. Many households on the "wrong" side of the digital divide are likely to use mobile devices and plans as their primary method of internet access, given lack of alternatives due to availability and cost. Thus, in the first chapter, we examine whether improving mobile internet access is an effective method for closing the digital divide. we do this by studying the adoption and use of unlimited mobile data plans offered by a large telecommunications provider. We find that adoption of an unlimited plan leads to a substantial increase in a household's data consumption, with the increase being particularly large for rural households and those of low socioeconomic status. This suggests that unlimited plans help these households "catch up", potentially narrowing the digital divide. We also study how households are using the additional data. Although most of the increase is accounted for by media and entertainment content, there is a significant increase in consumption of content likely to be socially beneficial: specifically news, education, and career-related content. We conclude that policy makers should encourage unlimited mobile data plans as a method to close the digital divide. They should also invest in educational programs on how to use the internet for beneficial purposes and help web site providers make their services accessible via mobile phones.

The emergence of the maker movement has democratized the production of goods and increased consumer demand for designer goods, leading to the founding of e-commerce platforms such as Etsy, Fab, and Gilt. Designers with little brand recognition have to overcome high buyer search costs in order to succeed in these markets or on their own websites. The second chapter examines two mechanisms that designers use to promote their products—flash sales and social media. Using two different identification techniques, we find that social media activities such as Facebook Likes, Pinterest Pins, and Faves have positive but different magnitudes of causal impact on sales, moderated by product type. Also, flash sale promotions increase average daily sales on the designer's primary website by up to 0.8 units in the first week after a flash sale is initiated. Overall, we find social media can be a strategic tool for designers to understand what kinds of products will succeed, promote products to compete against established brands like Amazon.

The third chapter examines how promotion of free Wi-Fi hotspots impacts both paid mobile and free Wi-Fi data usage. One could argue that promoting free Wi-Fi might cannibalize usage of paid mobile data. At the same time, increased usage of Wi-Fi might prime users into consuming more data, which might increase usage of mobile data. Interestingly, we find that promoting Wi-Fi hotspots leads to a small but significant mobile data usage, with the heavy mobile users having the highest impact. Also, increase in Wi-Fi usage is moderated by the type of business (or location) that provides the hotspots. Public places like airports, convention centers where people spend a lot of time have the highest increase in Wi-fi usage. Overall, our study reveals how and where Wi-Fi last mile channel can complement mobile internet usage.

These studies indicate that technology such as mobile internet, e-commerce platforms etc. are tools that have the potential to create a level playing field for all sections of the society. However, policy makers, managers and individuals must understand the mechanisms at play so that they can design policies to yield better outcomes.

1.1 References

Atasoy H (2013) The effects of broadband internet expansion on labor market outcomes. ILR Review 66(2):315–345.

Bakos JY (1997) Reducing Buyer Search Costs: Implications for Electronic Marketplaces. Management Science 43(12):1676–1692.

Bonfadelli H (2002) The Internet and knowledge gaps: A theoretical and empirical investigation. European Journal of Communication 17(1):65–84.

Christensen G, Steinmetz A, Alcorn B, Bennett A, Woods D, Emanuel EJ (2013) The MOOC phenomenon: Who takes massive open online courses and why? SSRN (November 6) https://papers.ssrn.com/abstract=2350964.

DiMaggio P, Hargittai E, Celeste C, Shafer S (2004) From unequal access to differentiated use: A literature review and agenda for research on digital inequality. Neckerman K, ed. Social Inequality. (Russell Sage Foundation, New York).

Forman C, Goldfarb A, Greenstein S (2005) How did location affect adoption of the commercial Internet? Global village vs. urban leadership. Journal of Urban Economics 58(3):389–420.

Greenstein S, McDevitt RC (2011) The broadband bonus: Estimating broadband Internet's economic value. Telecommunications Policy 35(7):617–632.

Hargittai E (2008) The Digital Reproduction of Inequality.

Nattamai Kannan KB, Hu YJ, Narasimhan S (2016) Social Media, Flash Sales, and the Maker Movement: An Empirical Analysis (Social Science Research Network, Rochester, NY).

Malone TW, Yates J, Benjamin RI (1987) Electronic Markets and Electronic Hierarchies. Communications of the ACM 30(6):484–497.

Prior M (2005) News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout. American Journal of Political Science 49(3):577–592.

Tichenor PJ, Donohue GA, Olien CN (1970) Mass media flow and differential growth in knowledge. The Public Opinion Quarterly 34(2):159–170.

Weber M (2013) Economy and Society Reprint edition. Roth G, Wittich C, eds. (University of California Press, Berkeley).

Zook MA (2000) The web of production: the economic geography of commercial Internet content production in the United States. Environment and planning A 32(3):411–426.

Zook MA (2001) Old hierarchies or new networks of centrality? The global geography of the Internet content market. American Behavioral Scientist 44(10):1679–1696.

CHAPTER 2. CAN THE MOBILE INTERNET BRIDGE THE DIGITAL DIVIDE? A LARGE-SCALE EMPIRICAL INVESTIGATION

2.1 Introduction

"….one of my top priorities would be to close the digital divide—the gap between 'those who use cutting-edge communications services and those who do not'…."

Ajit Pai, Chairman of the Federal Communications Commission (Pai 2017)

Access to the internet has become a de facto requirement for participating in contemporary society. People without good internet access may be unable to apply for jobs, they may have difficulty completing school assignments, and they may miss out on broad swaths of contemporary culture (Agarwal et al. 2009, Ransbotham et al. 2016). Unfortunately, there are many people in the United States (as well as around the world) who lack good internet access, particularly those with low socioeconomic status (SES) or who live in rural areas (DiMaggio et al. 2004, Greenstein and Prince 2009). Some have no access, while others are limited by poor connectivity or usage restrictions such as monthly data caps. Thus, those who might benefit the most from good internet access – and its ability to provide access to jobs, education, and knowledge – may be the least likely to have it. This "digital divide" is well-documented (Monica and Andrew 2016) and illustrated by a recent anecdote reported by Talbot (2016) about a 13-year-old girl living in a public housing project in Cleveland, Ohio. Her special-education plan requires her to complete math problems online using Khan Academy. However, she has no home broadband internet access and accessing the Khan Academy lessons through her mobile phone would quickly exhaust her family's data cap.

Finding ways to close the digital divide – and to allow more people to participate fully in our increasingly digital society – is a significant public policy issue. One solution is to build new telecommunications infrastructure so that more people have access to broadband internet service. However, building fixed telecommunications infrastructure (e.g., installing fiber-optic cables) is an expensive, slow, and labor-intensive process. For example, Google has suspended installation of fiber-optic cables for its broadband internet service because of high costs (Brodkin 2016). A better solution may be to leverage mobile technologies. A single mobile tower can provide connectivity to thousands of users in a given geographic area. Given that smartphone penetration exceeds 80%, mobile technology represents a relatively low-cost access option (Monica and Horrigan 2016). Indeed, given the likely cost advantage of mobile over fixed connections, the Federal Communications Commission (F.C.C.) has created a $4.5 billion fund to advance mobile broadband service, primarily in underserved rural areas (F.C.C. 2017a). Despite this initiative, there is limited empirical evidence of whether or how providing better mobile broadband internet access will help close the digital divide. We address this gap by studying the phased adoption of unlimited mobile data plans offered by a large telecommunications provider. We use a quasi-experimental research design to examine what happens when households switch from a metered to an unrestricted mobile data plan, which are becoming increasingly popular and are available from all major telecom providers in the U.S. Using a difference-in-differences approach, we find that adoption of unlimited data plans leads to increased data use on average, with this increase being particularly large for households on the "wrong" side of the digital divide, specifically, households in rural areas and those with low socioeconomic status. This suggests that mobile telecommunications technology, coupled with the right service plans from telecommunications providers, can help close the digital divide.

To understand the public policy implications of this finding, it is important to understand not only how much additional data is consumed, but also what this data is used for. For example, if the additional data is used predominantly for entertainment purposes (such as watching cat videos on YouTube), then closing the digital divide might not yield the public welfare benefits that many advocates expect. It is also possible that the additional data could exacerbate social inequality. Such as outcome would be consistent with the knowledge gap hypothesis (Tichenor et al. 1970), which states that as more information is made available to a social system, segments with high socioeconomic status will assimilate it faster than segments with low socioeconomic status, such that the gap in knowledge (and the corresponding social advantages) tends to increase. On the other hand, if the additional data is used to access news, education, and career-related information, then closing the digital divide might generate social benefits such as improved civic engagement as well as better education and job prospects for traditionally disadvantaged populations. To examine this, we analyze data on which web sites households accessed after adopting an unlimited plan, which we categorize into nine categories: media, social, technology, sports, business, shopping, news, education, and career. We find that households who experience the largest increases in data consumption – rural households and urban households with low socioeconomic status – increase their consumption of data in all categories. This includes an approximately 25% increase in consumption of news and education content. Although this increase is smaller than the increase in media / entertainment consumption (which comprises the bulk of the increase), it suggests that adoption of unlimited mobile data plans leads to increased consumption of information likely to be welfare enhancing. We find similar results when we examine the increase in the number of sessions instead of the increase in data consumption, given that some types of content may be more data-intensive than others.

Our analysis contributes to the literature on the digital divide and has public policy implications. We show that unlimited mobile data plans are an effective tool for policy makers seeking to close the digital divide, and the particular pattern of our results allows us to comment on why. We believe the large gains for rural households are because they often lack access to high-quality fixed internet service, given a lack of infrastructure in their communities. As a result, the mobile internet may be their main method of connectivity, such that removal of a data cap results in particularly large increases in use. The same mechanism may also account for the increase in use by urban households with low socioeconomic status. As with rural households, these households may not have access to high-quality fixed internet service at their residences, given the lack of incentives for telecommunications providers to provide fixed connections to these households. Another likely mechanism is cost, because those low-SES urban households who have access to fixed internet service may have trouble paying for it. Difficulty accessing fixed internet service could make the mobile internet the primary method of connectivity for these households, such that removing usage restrictions results in large gains. In addition, we find evidence that the increased data consumption is welfare enhancing, although perhaps not to the degree that many advocates would hope. Thus, policy makers should continue to invest in educational programs on how to use the internet for socially beneficial purposes. Last, although unlimited mobile data plans can help traditionally disadvantaged households bridge the digital divide, much of the increased data they consume is via mobile phone. The small form factor of a mobile phone may make it difficult for users to perform complex tasks such as submit a job application or participate in an online course (Monica and Horrigan 2016). Thus, policy makers can incentivize producers of educational, news, and career-related content to make their services accessible via mobile phones. This is consistent with research that suggests that making internet content more usable can help close the digital divide (Viard and Economides 2014).

2.2 Literature Review

We contribute to two related streams of research: 1) the digital divide and digital inequality, and 2) the knowledge gap hypothesis

2.2.1 The Digital Divide and Digital Inequality

The digital divide or digital inequality refers to differences across groups in the adoption and use of information and communication technology (ICT) (DiMaggio et al. 2004). Put simply, a digital divide (or inequality) exists when some groups can easily access and/or use ICT, while other groups struggle to access and/or use ICT (Racherla and Mandviwalla 2013). In this paper, we focus on inequality in internet use. Two key factors that contribute to digital inequality in internet use are geography (or location) and socioeconomic status. Historically, households in rural areas and/or of low socioeconomic status have relatively limited access to the internet and to ICT in general, placing them on the "wrong" side of the digital divide. This is often referred to as the "geographical digital divide" and the "socioeconomic digital divide". A key reason for the geographical divide is that telecommunications providers have greater economic incentives to build the physical infrastructure for internet access in urban areas than in rural areas, given population density and economies of scale (Greenstein 2005, Greenstein 2015, Greenstein and Prince 2009, Levitz and Bauerlein 2017). Simply put, an infrastructure investment in an urban area is more likely to pay off because the provider can use it to serve more customers. The socioeconomic divide exists for similar reasons: individuals with low incomes are less likely to have high-quality internet access, perhaps due to lack of access at their residences and/or inability to pay (Caumont 2013, Compaine 2001, Dutton et al. 1989).

Given that the internet has become an integral part of modern society and a driving force of the global economy, closing the digital divide and reducing digital inequality is important for private enterprises, policy makers, and society in general. Recognizing this, both policy makers and telecommunications providers have implemented programs designed to reduce digital inequality. A central tenet of U.S. telecommunications policy has been the "universal service" principle that every American have access to telephone service. This principle helped dial-up internet access diffuse rapidly across the U.S., as it piggy-backed on the existing telephone infrastructure that the F.C.C. had promoted through a series of programs over the years. However, the next generations of internet access – broadband and wireless – required installation of new infrastructure. Telecommunications providers tended to install this infrastructure in relatively affluent (and thereby profitable) areas, creating digital inequality. The F.C.C. has implemented policies to reduce this inequality, such as subsidizing internet access (F.C.C. 2017b, Kang 2016, Ruiz 2015) and mandating that telecommunications providers provide internet access to less profitable areas to gain approval for mergers and acquisitions (AT&T 2015, Meg 2016). While policy makers have focused on incentives and mandates, telecommunication providers have focused on improving their technology and infrastructure to provide both cost-effective and high-speed internet service to the masses (Talbot 2016a).

Several papers have examined how governmental regulation, policy, and technological innovations influence the digital divide and digital inequality. As discussed by Greenstein (2015), governmental regulation and programs have shaped the evolution of the internet – and the resulting digital inequality – throughout its history. The Council of Economic Advisors (2015, p. 9) summarized how U.S. policy aims to close the digital divide by "including infrastructure investments and robust competition policy to ensure widespread access to affordable high-quality Internet; spectrum policy to ensure that the dramatic growth in wireless broadband continues; and investments in education and training, especially for children, to remove computer literacy barriers that impede universal access." Prior studies in the IS literature have examined governmental initiatives designed to close the digital divide. For example, Hsieh et al. (2008, 2010) examined a governmental program to provide free internet access in a mid-size U.S. city, focusing on differences in use intentions between socioeconomic groups. Venkatesh and Sykes (2012) examined a governmental program to provide internet access via kiosks in a rural village in India, finding that social network constructs predict both internet use and downstream economic benefits.

Technological changes – particularly the introduction of broadband service – have also led to positive outcomes and helped to close the digital divide for rural households and those of low socioeconomic status. Multiple studies have used aggregate data to show that access to broadband or broadband expansion leads to positive economic outcomes, particularly for rural areas (Atasoy 2013, Kolko 2012). Goldfarb and Prince (2008) examined individual-level survey data and found that, conditional on adoption of broadband, low-income individuals spend more time online compared to high-income individuals, which they attributed to the opportunity costs of leisure time. This suggests that broadband may help individuals of low socioeconomic status close the divide, although Kolko (2010) used individual-level internet subscription data to show that although broadband adoption increased overall internet use, there was no significant increase in "socially desirable" categories such as jobs and career. Hitt and Tambe (2007) examined the effect of switching from dial-up to broadband connections for a panel of households. They found that adoption of broadband led to increased data usage and a more even distribution of use among adopters, which suggests a reduction in digital inequality. We extend these studies by shifting the focus from fixed internet service to mobile internet service. Because the mobile internet is different from the PC-based internet (Ghose et al. 2012), the effects of the mobile internet on digital inequality may also be different. Also, the lower cost of providing internet access via mobile infrastructure vs. fixed infrastructure makes it important to study whether mobile internet access can help close the digital divide. Indeed, Prieger (2013) noted the potential of mobile broadband to benefit traditionally underserved rural areas. Last, Dewan et al. (2010) showed that adoption of PCs helps close the digital divide by providing a means of fixed internet access. It is important to study whether the rapid adoption of smartphones – which provide a means of mobile internet access – may have a similar effect on digital inequality.

2.2.2 The Knowledge Gap Hypothesis

The knowledge gap hypothesis is closely related to the socioeconomic digital divide. First proposed by Tichenor et al. (1970), the knowledge gap hypothesis is that "as the infusion of mass media information into a social system increases, segments of the population with higher socioeconomic status tend to acquire this information at a faster rate than the lower status segments, so that the gap in knowledge between these segments tends to increase rather than decrease." The knowledge gap hypothesis does not mean that people with low socioeconomic status will not gain useful information. Instead, it emphasizes that the gap in knowledge between those with high socioeconomic status and those with low socioeconomic status will grow over time.

Early tests of the knowledge gap hypothesis focused on information provided by newspapers and television. Bonfadelli (2002) was among the first to apply the knowledge gap hypothesis to information provided by the internet. He found support for the hypothesis by showing that users with higher levels of education tended to use the internet for gaining information while users with lower levels of education tended to use it predominantly for entertainment. Hargittai and Hinnant (2008) reached a similar conclusion in their study of young adults who use the internet. They found that those with higher levels of education and a more resource-rich background were more likely to use the internet for "capital enhancing" activities. Several more recent studies have shown that expansion of broadband internet access improves economic and social outcomes as well as civic engagement (Ganju et al. 2015, Whitacre et al. 2014, Whitacre 2017, Whitacre and Manlove 2016). However, the benefits are perhaps not as great as advocates of universal internet access would hope. For example, massively online open courses (MOOCS) exploded in popularity in recent years with the promise of making high-quality courses available to the world via the internet. However, evidence indicates that such courses are serving well-educated, employed individuals mostly in developed countries, instead of making a difference in the lives of those who have traditionally lacked access to higher education (Christensen et al. 2013). In a related finding, Helsper and van Deursen (2015) concluded that the benefits of using the internet are higher for those with higher socioeconomic status, such that offering more services online may exacerbate social inequality. These studies illustrate that the internet is merely a tool. Whether an individual or society benefits from it depends on how it is used (Wei et al. 2010). This illustrates the importance of our investigation of the types of content that households consume after adopting an unlimited data plan, which can help determine whether policy interventions are necessary to promote use of the internet for socially beneficial purposes.

2.3 Data and Empirical Strategy

2.3.1 Empirical Context and Data Overview

To examine whether the mobile internet can bridge the digital divide, we analyze detailed subscriber-level data provided by a large U.S. telecommunications provider[1] (hereafter referred to as the partner or company). The company is a leading provider of a full range of telecommunication services, including wireless voice and data service and TV service for most parts of the U.S. In January 2016, the company announced a new program that allowed existing and new customers to subscribe to an unlimited mobile data plan.[2] The data plan provides unlimited high speed mobile internet access[3] to all individuals subscribed via the same mobile account (e.g., a family's account might cover multiple family members). Prior to introduction of this plan, the company had primarily offered data plans with a fixed data limit. The data spans from January 2015 to December 2016 and contains detailed information on mobile data use for each subscriber. We measure data use at the household level, which is the sum of individual data use of all subscribers in that

---

[1] The company wishes to remain anonymous.

[2] To continue with the unlimited plan after the first 3 months, the subscriber had to subscribe to at least one additional service, such as TV.

[3] The partner reserved the right to throttle speeds after a subscriber reaches certain data limits, if there is congestion in the network at that time.

household. We measure data use for every calendar month to create a household/month panel from January 2015 to December 2016. The data contains not only aggregate usage information but also information on how the data was used. The company categorizes data use into nine different categories based on which web site or service the household uses. The data also contains demographic information about each household, including income and geographic location.

Working with the company allows us to overcome several challenges that have limited researchers' ability to study whether the mobile internet can bridge the digital divide. First, we have household-level on actual data consumption (including the type of content consumed), which is difficult if not impossible to obtain without collaborating with a telecommunications provider. Second, we have data for a large pool of households, which allows us to examine the effect of adoption of the unlimited plan for households on both sides of the digital divide. Third, we have data before and after the introduction of the unlimited plan, both for households that adopted the plan and those that did not. This provides us with an identification strategy to examine the likely causal impact of the unlimited plan (which we describe below).

We restrict the sample in three ways. First, we include only non-commercial accounts, given our interest in studying the effect of the mobile internet for households. Second, we include only households who were subscribers prior to January 2015 and who continued to be subscribers after December 2016. Third, we drop households that had multiple accounts or that switched addresses in the period under study. This allowed us to analyze stable households whose data use could be measured accurately and whose location (i.e., urban or rural) did not change.

2.3.2 *Identification Strategy*

The telecommunications provider's introduction of the unlimited mobile data plan in January 2016 created a quasi-natural experiment with non-random assignment of treatment (Shadish et al. 2001, p. 15): some households adopted the new plan while others did not. Our basic strategy is to compare data use of the households who adopted (i.e., the treated households) to those that did not (i.e., the control households). To account for the non-random assignment of treated and control households, we use a difference-in-differences strategy coupled with coarsened exact matching (Iacus et al. 2012). This allows us to create a matched sample in which the control group is similar to the treated group, such that any increase in data use for the treated group can be attributed to adoption of the unlimited data plan rather than to other factors.

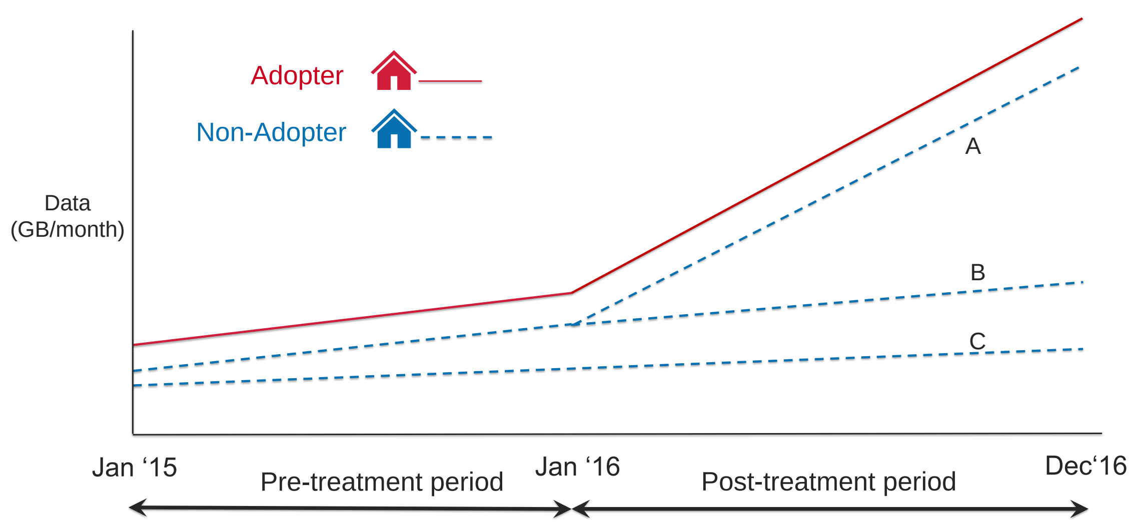

Figure 2-1 Illustration of research design.

Figure 2-1 illustrates our research design by showing the hypothetical monthly data use of an adopting (treated) household that adopted in January 2016. In the pre-treatment period (i.e., January 2015 – December 2015), the figure shows a gradual increase in the adopting household's data use. After adoption in January 2016, the household's data use increases rapidly and continues to increase throughout the post-treatment period. However, we need to determine whether this increase would have happened anyway – even if the household had not adopted the unlimited plan. Because we only observe the increase in data use after adoption, we have the classic 'missing data' problem encountered in the causal inference literature (Angrist and Pischke 2008). To overcome this, we need to identify a non-adopting household to serve as a valid counterfactual, i.e., what the trajectory of the adopting household's data use would have been had it not adopted the unlimited plan. Figure 1 illustrates multiple possibilities. The (A), (B), and (C) lines show the monthly data use of hypothetical non-adopting households. The (A) and (B) lines show a non-adopting household whose data use in the pre-treatment period is similar (in fact, parallel) to that of the adopting household. This similarity suggests that this household's use of data – including how that changes month-over-month – is similar to that of the adopting household, at least until the adopting household adopts. If the non-adopting household's data use follows a similar trajectory in the post-treatment period as in the pre-treatment period, as shown in (B), then the increase in data use for the adopting household is likely caused by adoption of the unlimited plan. On the other hand, if the trajectory of the non-adopting household increases in January 2016 along with that of the adopting household, as shown in (A), then this increase is likely caused by some unobserved change that affected both adopters and non-adopters – and not by adoption of the unlimited plan. The (C) line depicts another possibility. The pre-treatment trend for this non-adopting household is flatter than that for the adopting household. This would not be a good counterfactual, because an observed increase in data use for the adopting household in

January 2016 could reflect the effect of the unlimited plan or a continuation of the differential pre-treatment trend (or both).

2.3.3 Constructing the Matched Sample

We used coarsened exact matching to create a matched sample in which each adopting household has a valid counterfactual. We did this by matching adopting (treated) households to non-adopting (control) households on data use in the pre-treatment period, socioeconomic status, and geographic location.

2.3.3.1 Matching Procedure: Data Use in the Pre-Treatment Period

To illustrate how we matched on data use in the pre-treatment period, let us first consider households that adopted the unlimited plan in January 2016. We matched these treated households to control households based on the following variables: a) actual data used in November 2015 and December 2015, and b) the change in data used between the following months compared to the prior month, all in 2015: April, June, August, October, and December. We coarsened each variable into one of 20 distinct bins and only matched treated and control households whose values were in the same bins. For example, assume that a treated household used 7 GB of data in November and 8 GB in December and increased its data use by 0.5 GB between March and April, then again between May and June, etc. We matched this treated household to a control household with similar values for each variable (specifically, whose values were in the same bin). This ensured that treated and control households had similar data use trends in the pre-treatment period. Matching on actual data use ensures that matched households have a similar magnitude of data use, and matching on changes in data use ensures that matched households have a similar trajectory of data use. We referred to these treated and control households as the “January cohort”. We then set aside this cohort so that they were not used in the next step. Next, consider households that adopted the unlimited plan in February 2016. We used the same procedure to find matching control households, except after incrementing the months used for the matching. For example, instead of using the actual data use for November and December, we used the actual data use for December and January. These treated and control households are the “February cohort”. We proceeded similarly with households that adopted in the other months in 2016.[4] This yielded a total of 12 cohorts. Figure 2-2 shows monthly data use for households that adopted the unlimited plan in January, February, March, and April of 2016.

[4] We implemented this “rolling” matching procedure using R and plan to release the code as an R package. We believe this will be helpful for other researchers who need to conduct matching when treated subjects are treated at different times.

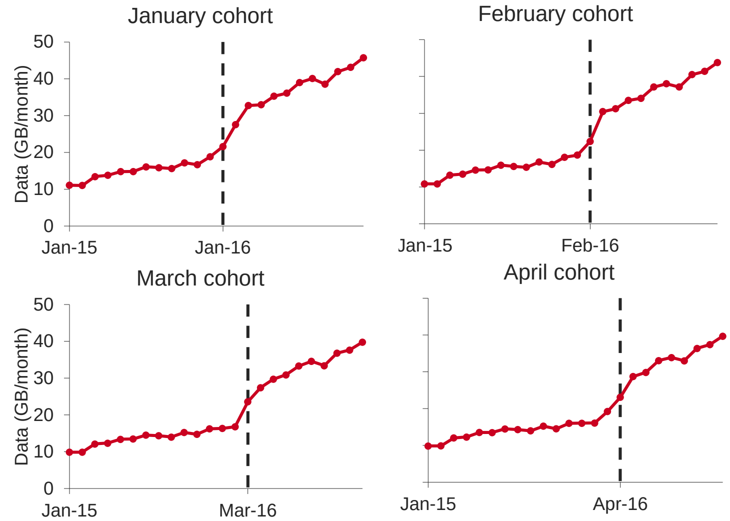

Figure 2-2 Average data use of households who adopted an unlimited plan in January, February, March, or April

2.3.3.2 Description of the Matched Sample

We dropped treated households for which we could not find a match. In cases in which we found more than one match for a treated household, we randomly selected one of the control households. Thus, each treated household has a single match. There are 1,080,158 households in the matched sample: half are treated and half are controls. Because we observe each household for 24 months, our panel data set contains 25,923,792 observations.

Table 2-1 shows the distribution of the matched sample across socioeconomic and geographic groups. Table 1 shows that even the smallest cell – Mostly Rural-Low SES – has almost 3,000 observations. As such, we believe that we have enough data to generate valid conclusions for each socioeconomic / geographic combination.

Table 2-1 - Distribution of households in the matched sample by socioeconomic and geographic group

	Urban	Mostly Urban	Mostly Rural	Rural	**Total**
Low SES	9,852	15,936	2,998	5,702	34,488
Mid SES	249,428	356,690	53,426	60,232	719,776
High SES	135,338	175,764	9,628	5,164	325,894
Total	394,618	548,390	66,052	71,098	**1,080,158**

Table 2-2 Summary statistics for variables in the matched sample lists variables and shows summary statistics for the matched sample. $Unlimited_{it}$ is a dummy variable that takes a value of 1 for the adopting households after they adopted and 0 otherwise. $Data_{it}$ measures the monthly data use per household in GB/month (at times, we also measure $Data_{it}$ in MB/month, as explained below). The correlation between $Unlimited_{it}$ and $Data_{it}$ is 0.39, providing model-free evidence that adoption leads to an increase in data use. The near zero correlation between $Unlimited_{it}$ and $Percentage\ Rural_i$ and $Income_i$ gives us confidence that there is no meaningful difference between the treated and control households for these variables in the matched sample.

Table 2-2 Summary statistics for variables in the matched sample

	Variable	Mean	Std. Dev.	Min	Max	(1)	(2)	(3)
1	$Unlimited_{it}$	0.14	0.35	0.00	1.00	1.00		
2	$Data_{it}$ (GB/month)	17.74	21.04	0.00	2206.07	0.39***	1.00	
3	$Income_i$	3.66	1.15	1.00	5.00	0.02***	0.00**	1.00
4	Percentage $Rural_i$	0.18	0.29	0.00	1.00	0.00***	0.06***	-0.19***

Notes: These variables are measured for 1,080,158 households over 24 months, yielding a total of 25,923,792 observations. In our main analysis, $Percentage\ Rural_i$ is coarsened into four bins: Urban, Mostly Urban, Mostly Rural, and Rural. $Income_i$, measured as of December 2015, is an ordinal variable measuring the household monthly income with five categories: 1 - Under \$19,999; 2 - \$20,000-\$39,999; 3 - \$40,000-\$74,999; 4 - \$75,000-\$124,999; 5 - \$125,000 and over. In our main analysis, we coarsen these into 3 bins: Low SES, Mid SES, and High SES.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

As noted above, the company that provided the data categorizes the web sites and services that households visit into one of nine categories for its own reporting purposes.

Table 3 describes each category, including sample web sites. Table 2-4 shows summary statistics for $Data_{it}$ by category. Because some web sites, such as video-heavy sites such as Netflix, are data-heavy compared to others, we also measure the number of sessions per category ($Sessions_{it}$). Once a household begins a session with a web site (say netflix.com), the company considers data used in the next 15 minutes to be for that session. If the household continues using the web site beyond the first 15 minutes, the company records that as a second session. A third session is recorded if use continues for another 15 minutes, and so on.

Not all of the data used by subscribers is for access to web sites and services. Our main $Data_{it}$ variable captures the total data transferred between the cell tower and the mobile phone, which we refer to as network-level data. A subset of this data is used when a household accesses a specific web site or service. We refer to this as category-level data. By definition, only the category-level data is grouped into the categories shown in Table 2-3. Also, due to the company's data storage procedures, category-level data was only available for a representative sample of households from November 2015 to October 2016. Later in the paper, we show that the category-level data is representative of the network-level data.

Table 2-3 Description of content categories

Category	Examples	Description
Media	netflix.com, pandora.com	Web sites for audio and video entertainment.
Social	facebook.com, instagram.com	Web sites designed for social interaction.
Technology	google.com, amazonaws.com	Search engines, email services, cloud-based file sharing services, etc.
Sports	mlb.com, nfl.com	Web sites dedicated to sports and online gaming.
Business	doubleclick.net, moat.com	Primarily third-party web sites that serve ads.
Shopping	amazon.com, bestbuy.com	Web sites that sell products and services.
News	nytimes.com, cnn.com	Web sites of newspapers, magazines, etc.
Education	academia.edu, haskell.edu	Web sites of educational institutions, sites for sharing academic material, etc.
Careers	glassdoor.com, careerbuilder.com	Job posting and recruiting web sites.

Table 2-4 Summary statistics for data use by category

Category	$Data_{it}$ (MB/Month)				$Sessions_{it}$ (Count/Month)			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Media	2,386.10	6,369.39	0	437,186.63	1,243.62	1,589.16	0	44,831.00
Social	1,466.33	2,424.43	0	136,834.22	1,834.35	2,544.13	0	237,903.00
Technology	1,363.75	2,448.06	0	1,090,000.00	6,687.35	6,832.12	0	368,425.00
Sports	80.44	506.51	0	95,340.20	227.22	591.66	0	41,224.00
Business	419.42	902.56	0	106,961.91	5,214.37	5,711.48	0	339,853.00
Shopping	50.25	132.15	0	51,979.70	295.97	472.24	0	49,145.00
News	44.85	158.5	0	50,922.29	224.83	419.47	0	30,522.00
Careers	1.49	12.51	0	4,928.63	10.93	46.85	0	6,284.00
Education	22.16	176.15	0	134,693.75	60.36	221.85	0	14,695.00
Pooled	5,834.80	9,376.00	0	1,090,000.00	15,798.98	15,912.62	0	596,879.00

2.3.4 Regression Specifications

We use a difference-in-differences (DID) approach to estimate whether and how households changed their data use after adopting the unlimited data plan. This approach has been used extensively by scholars in economics, information systems, and other fields

(Bertrand et al. 2004, Burtch et al. 2018, Greenwood and Agarwal 2015, Greenwood and Wattal 2017). The DID regression specification is shown in model (1) below:

$$Y_{it} = \alpha + \beta\, Unlimited_{it} + \sum_{k=1}^{12}\sum_{p=1}^{12}(T_t * Cohort_p * Geosocial_k) + h_i + \varepsilon_{it} \tag{1}$$

In (1), Y_{it} is the data used by household i in month t, which is either $Data_{it}$ or $\log(1+Data_{it})$. When we analyze the category-level data, Y_{it} is monthly data use for a given category. As described above, $Unlimited_{it}$ is a dummy variable equal to 1 if household i had the unlimited data plan in month t. $Cohort_p$ are dummy variables that reflect a household's cohort (e.g., the January cohort, the February cohort, etc.). $Geosocial_k$ are dummy variables that reflect a household's geosocial group. α is the constant term, T_t are month fixed effects, h_i are household fixed effects, and ε_{it} is the error term. The household fixed effects capture all time-invariant household-specific variables such as demographics, geographic location, etc. We interact the month fixed effects with $Cohort_p$ and $Geosocial_k$ to allow the time trends for the 24 months of our sample to differ for each of the 144 combinations of $Cohort_p$ and $Geosocial_k$.

We extend model (1) by interacting $Unlimited_{it}$ with the $Geosocial_k$ dummy variables as well as dummy variables for geographic location and socioeconomic status separately. This allows us to examine how the treatment effect of the unlimited plan varies across groups.

A key test of the validity of our DID model is whether the pre-treatment trends in data use are parallel for the treated and control households (Bertrand et al. 2004). Parallel pre-treatment trends increase the likelihood that a change in data use following adoption of the unlimited plan is, in fact, due to adoption of the unlimited plan. We test for parallel pre-treatment trends using a relative time model or leads/lags model (Greenwood and Agarwal 2015), as specified in model 2 shown below:

$$Y_{it} = \alpha + \sum_{\tau=-23}^{\tau=-2} \rho_\tau Unlimited_{i(t+\tau)} + \sum_{\tau=0}^{\tau=11} \rho_\tau Unlimited_{i(t+\tau)} + \sum_{k=1}^{12}\sum_{p=1}^{12}(T_t * Cohort_p * Geosocial_k) + h_i + \varepsilon_{it} \quad (2)$$

Specification (2) is exactly same as (1) except that we replace the term $\beta\ Unlimited_{i,t}$ with $\sum_{\tau=-23}^{\tau=-2} \rho_\tau Unlimited_{i(t+\tau)} + \sum_{\tau=0}^{\tau=11} \rho_\tau Unlimited_{i(t+\tau)}$. $Unlimited_{i(t+\tau)}$ is a dummy variable equal to 1 for observations in month *t* if month *t* is τ months after adoption of unlimited plan (or for $\tau < 0$, $-\tau$ months before adoption of unlimited plan). These dummies range from -23 for those who adopted in December 2016 to +11 for those who adopted in January 2016. We withhold $\tau = -1$ to avoid the dummy variable trap. This means that the month before adoption is used as the baseline. Because $Unlimited_{i(t+\tau)}$ is always 0 for the control households, the coefficient $\rho_{-\tau}$ for $Unlimited_{i(t-\tau)}$ represents the average difference in total monthly data use between the treated and control households τ months before the month of adoption for the treated households. If pre-treatment trends are parallel, then the coefficients for $Unlimited_{i(t-2)}$, $Unlimited_{i(t-3)}$, etc. will be both economically and statistically insignificant. The coefficients for $Unlimited_{i(t+1)}$, $Unlimited_{i(t+2)}$, etc. allow us to assess whether the treatment effect changes over time.

2.4 Results

2.4.1 Model-Free Evidence

Figure 2-3 plots the monthly data use of the January cohort, i.e., the treated households who adopted the unlimited data plan in January 2016 and the non-adopting control households matched to them. In the pre-treatment period from January to December 2015, monthly data use was essentially the same for both groups. Thus, the data appears to satisfy the parallel pre-treatment trends assumption for DID models. From January 2016 onward, there is a sharp increase in data use for adopting households compared to non-adopting households. This suggests that adoption of the unlimited data plan had a large effect on data use. One concern is that a different event in January 2016 – other than adoption of the unlimited data plan – might have caused this increase. However, this explanation would only be valid if the hypothetical event affected the treated households but not the control households. To further rule out this alternate explanation, we created monthly data use plots for the treated and control households in the other cohorts (February, March, etc.) In each instance, we found parallel pre-treatment trends followed by a sharp increase in data use for the treated households.

Figure 2-4 shows monthly data use plots for four of the geosocial groups in the January cohort: Urban-High SES, Urban-Low SES, Rural-High SES, and Rural-Low SES. The plots suggest that parallel pre-treatment trends hold for each of these groups, and that adoption has a larger influence for rural households and those with low socioeconomic status.

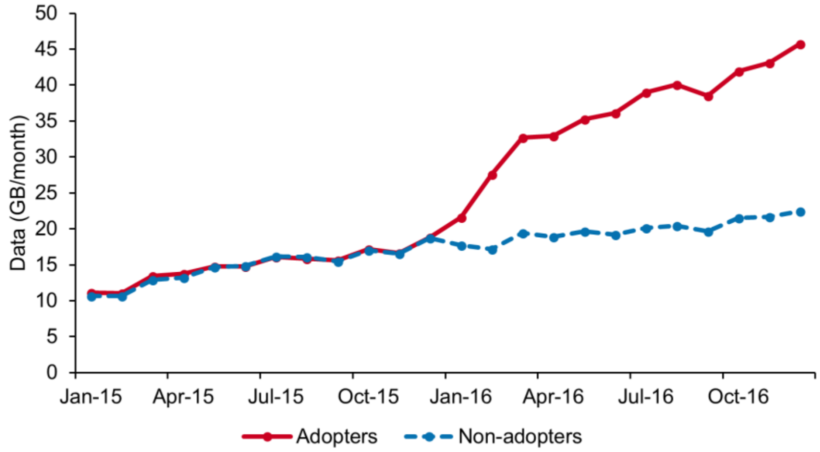

Figure 2-3 Average data use of adopters and matched non-adopters in the January cohort.

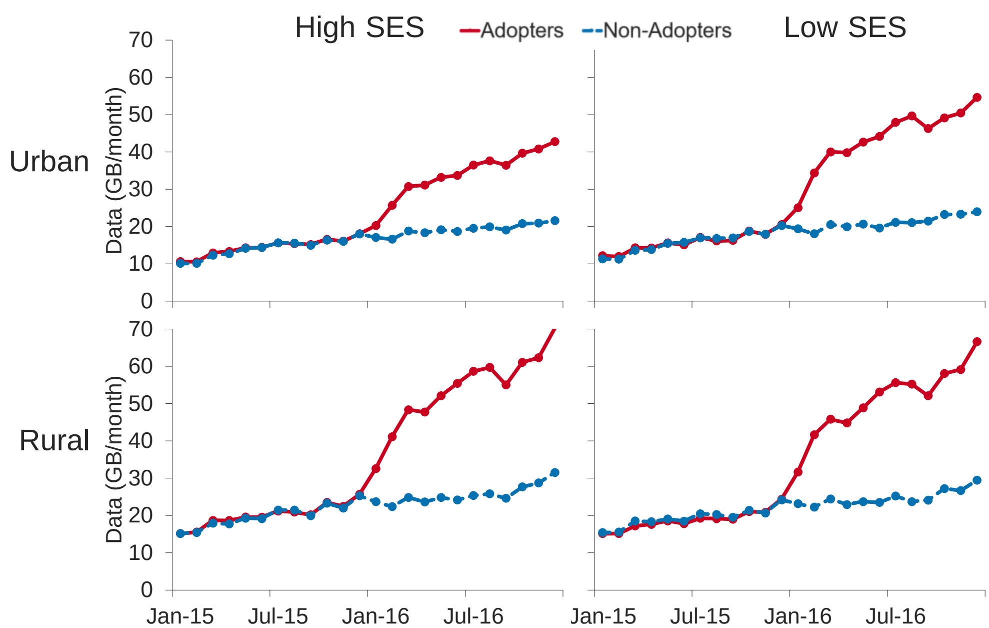

Figure 2-4 Average data use of adopters and matched non-adopters in the January cohort, broken down by and SES and geography

2.4.2 Regression Results: Overall Data Use

Column 1 of Table 5 shows the results of model (2), which allows us to test the parallel trends assumption for difference-in-differences estimation. Column 1 shows the results for the full matched sample. Columns 2-6 show the results when we limit the sample to the January, February, March, April, and May cohorts. In each case, the coefficients of the lagged terms ($Unlimited_{i,-5}$ to $Unlimited_{i,-2}$) are small and often insignificant. This indicates that there is no meaningful difference between treated and control households in the pre-treatment period, i.e., that the pre-treatment trends are (generally) parallel. By comparison, the lead terms ($Unlimited_{i,+0}$ to $Unlimited_{i,+3}$) are large, highly significant, and increasing over time. This indicates that adoption of the unlimited plan causes households to use more data, and increasingly so over time. Table 5 reports only a subset of the lagged and lead coefficients due to space constraints. Figure 2-5 shows all of the lagged and lead coefficients. The lagged coefficients are close to zero while the lead terms are increasingly positive over time.

Table 6 shows the results of model (1) with the dependent variable as either $Data_{it}$ or log(1+ $Data_{it}$). As shown in column 1, adoption of the unlimited plan increases data use by an average of 17.8 GB per month. Column 4 reports the result when using logged data use as the dependent variable and shows an increase of approximately 52%. Columns 2 and 5 show the results broken down by geographic location. The baseline group consists of Urban households, for whom the treatment effect is 15.2 GB/month (or approximately 46%). The treatment effects for Mostly Urban, Mostly Rural, and Rural households are 3.05, 7.58, and 8.00 GB/month larger than that for Urban households. Columns 3 and 6 show the results broken down by socioeconomic status. The baseline group consists of high

SES households, for whom the treatment effect is 15.3 GB/month. The treatment effect is larger for mid SES (15.3 + 3.3 = 18.6 GB/month) and low SES (15.3 + 6.93 = 22.23 GB/month) households. This indicates the effects are larger for rural households and those with low socioeconomic status.

Table 2-5 Regression results for the relative time model (specification (2))

Model	Dependent variable: $Data_{it}$ (1) Pooled (all cohorts)	(2) Jan cohort	(3) Feb cohort	(4) Mar cohort	(5) Apr cohort	(6) May cohort
$Unlimited_{i,-5}$	0.28***	-0.10	-0.11*	0.15**	0.08	0.14*
	(0.02)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
$Unlimited_{i,-4}$	0.28***	0.07	0.04	0.18***	0.12*	0.01
	(0.02)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
$Unlimited_{i,-3}$	0.25***	0.08	0.07	0.17***	-0.00	0.02
	(0.02)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
$Unlimited_{i,-2}$	-0.08***	-0.11*	-0.05	-0.07	-0.20***	-0.11*
	(0.02)	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)
$Unlimited_{i,-1}$	Baseline reference group					
$Unlimited_{i,0}$	6.73***	5.05***	6.31***	6.59***	6.33***	6.64***
	(0.02)	(0.06)	(0.07)	(0.08)	(0.08)	(0.08)
$Unlimited_{i,+1}$	13.69***	13.26***	13.84***	12.30***	12.76***	13.89***
	(0.04)	(0.10)	(0.11)	(0.10)	(0.12)	(0.13)
$Unlimited_{i,+2}$	15.72***	16.71***	15.17***	14.44***	15.26***	16.79***
	(0.04)	(0.12)	(0.11)	(0.12)	(0.14)	(0.15)
$Unlimited_{i,+3}$	17.14***	17.36***	16.79***	16.50***	17.90***	17.39***
	(0.05)	(0.12)	(0.12)	(0.13)	(0.16)	(0.16)
Constant	9.74***	10.84***	10.20***	9.38***	9.24***	9.16***
	(0.02)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Observations	25,923,792	3,272,496	3,297,360	2,851,152	2,280,432	2,143,872
R-squared	0.15	0.14	0.16	0.15	0.15	0.15
Number of households	1,080,158	136,354	137,390	118,798	95,018	89,328

Notes: To conserve space, this table only shows the coefficients for 4 lead and lag terms. Also, results for the June to December cohorts are not reported. For the complete table with 23 lag and 12 lead terms and all 12 cohorts, please contact the authors. Fixed effects for households and months (interacted with cohort and geosocial group) included. Clustered standard errors (by household) in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

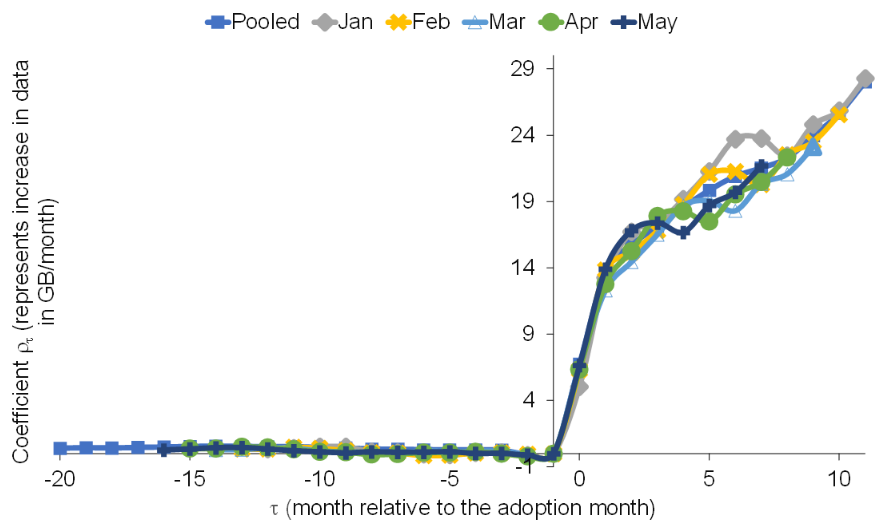

Figure 2-5 Plot of lead and lag coefficients for the relative time model (specification (2))

Table 2-6 Regression results for the main specification (specification (1))

	DV: $Data_{it}$			DV: Log(1+$Data_{it}$)		
	(1)	(2) Geography	(3) SES	(4)	(5) Geography	(6) SES
$Unlimited_{it}$	17.8*** (0.04)	15.2*** (0.06)	15.3*** (0.07)	0.52*** (0.00)	0.46*** (0.00)	0.45*** (0.00)
$Unlimited_{it}$ * Mostly Urban		3.05*** (0.09)			0.074*** (0.00)	
$Unlimited_{it}$ * Mostly Rural		7.58*** (0.21)			0.13*** (0.00)	
$Unlimited_{it}$ * Rural		8.00*** (0.22)			0.15*** (0.00)	
$Unlimited_{it}$ * Mid SES			3.30*** (0.09)			0.099*** (0.00)
$Unlimited_{it}$ * Low SES			6.93*** (0.30)			0.16*** (0.01)
Constant	22.6*** (0.02)	22.6*** (0.02)	22.6*** (0.02)	2.75*** (0.00)	2.75*** (0.00)	2.75*** (0.00)
R-squared	0.27	0.28	0.27	0.40	0.40	0.40

Notes: n = 25,923,792 observations (1,080,158 households x 24 months). Fixed effects for households and months (interacted with cohort and geosocial group) included. For models 2 and 5, the baseline reference group is Urban. For models 3 and 6, the baseline reference group is High SES. Clustered standard errors (by household) in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

Table A. 2 Regression results, including all 12 geosocial groups. in the appendix shows the regression results after including interaction terms to capture all 12 combinations of geography and socioeconomic groups: Urban-High SES, Urban-Mid SES, Urban-Low SES, etc. For simplicity and to conserve space, we focus on the 4 “extreme” groups: Urban-High SES, Urban-Low SES, Rural-High SES, and Rural-Low SES.

Table 2-7 shows the results after limiting the sample to households in these four groups. Columns 1 and 4 show that households in the Urban-High SES group increase their data use by 13.25 GB/month after adoption (or approximately 40%). Households in the Urban-Low SES, Rural-High SES, and Rural-Low SES group increase their use even more: by an additional 6.97, 9.99, and 10.60 GB/month, respectively. Figure 2-6 shows this result graphically. We re-use this visualization below to summarize the results of the category-level analysis. Column 2 and 5 show the results for the subset of urban households, while columns 3 and 6 show the results for rural households. Columns 2 and 3 show that there is a statistically significant difference in GB/month between low and high socioeconomic groups in urban areas but not in rural areas. However, column 6 shows that while the raw GB/month increase for rural households is similar for both high and low socioeconomic groups, the percentage increase is greater for households with low socioeconomic status (by approximately 9 percentage points, $p<0.01$).

Table 2-7 Regression results for the "extreme" SES and geographic groups

	DV: $Data_{it}$			Log(1+$Data_{it}$)		
Model	(1)	(2) Urban	(3) Rural	(4)	(5) Urban	(6) Rural
$Unlimited_{it}$	13.25***	13.25***	23.24***	0.40***	0.40***	0.54***
	(0.10)	(0.10)	(0.76)	(0.00)	(0.00)	(0.01)
$Unlimited_{it}$ * Urban-Low SES	6.97***	6.97***		0.17***	0.17***	
	(0.53)	(0.53)		(0.01)	(0.01)	
$Unlimited_{it}$ * Rural-High SES	9.99***			0.14***		
	(0.76)			(0.01)		
$Unlimited_{it}$ * Rural-Low SES	10.60***		0.61	0.23***		0.09***
	(0.76)		(1.07)	(0.01)		(0.02)
Constant	21.33***	20.66***	30.36***	2.73***	2.71***	2.99***
	(0.05)	(0.05)	(0.27)	(0.00)	(0.00)	(0.01)
Observations	3,745,344	3,484,560	260,784	3,745,344	3,484,560	260,784
R-squared	0.27	0.27	0.28	0.39	0.39	0.41
Number of households	156,056	145,190	10,866	156,056	145,190	10,866

Notes: Fixed effects for households and months (interacted with cohort and geosocial group) included. For models 2 and 5, the baseline reference group is Urban-High SES. For models 3 and 6, the baseline reference group is Rural-High SES. Clustered standard errors (by household) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

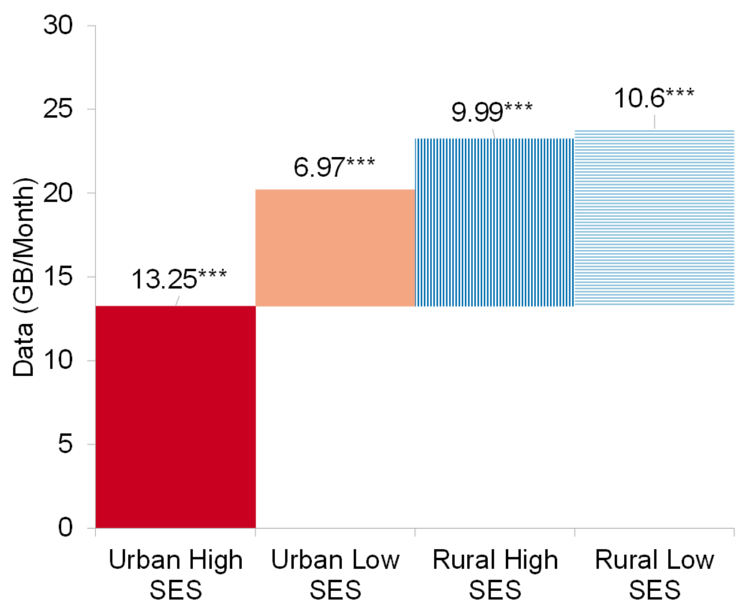

Figure 2-6 Visualization of the regression results for the "extreme" SES and geographic groups showing the average increase in data use for each group

2.4.3 Robustness Checks

In our main analysis, we clustered the standard errors by household. In addition to clustering the standard errors, we also block bootstrapped the standard errors to dampen the influence of outliers (Greenwood and Wattal 2017). Results using bootstrapped standard errors are similar to those in Tables 6 and 7.

Our results could be influenced by the breakpoints that we used to define the geographic and socioeconomic status categories. To examine this, we re-estimated the model after interacting the raw value of $Percentage\ Rural_i$ with $Unlimited_{it}$.. The results are shown in Table A. 1. As $Percentage\ Rural_i$ increases, so does the treatment effect. When $Percentage\ Rural_i$ is 1 (or 100%), the treatment effect is 8.8 GB/month higher than the baseline of 16.2 GB/month. This is consistent with our main findings. Similarly, we re-estimated the model after interacting the ordinal values of $Income_i$ noted in Table 2-2 with $Unlimited_{it}$ (we do not observe actual income, only an income range). These results are shown in Table A. 1 and are consistent with our main findings.

2.4.4 Regression Results: Data Use by Category

As noted in the previous section, we have data on category-level web site use (the category-level data) for a subset of our matched sample (the network-level data). To assess whether the category-level data was representative of the network-level data, we reran the regressions whose results appear in Table 2-7, but using the category-level data instead of the network-level data. We also report the coefficients in MB/month rather than GB/month (for reasons that we clarify below). As shown in Table 2-8, the pattern of effects using the category-level data is similar to that using the network-level data: the treatment effect is larger for rural households and those with low socioeconomic status.

Table 2-8 reports the results with overall data use as the dependent variable. Column 1 of Table 2-9, Panel A reproduces this result. Columns 2-10 of Table 2-9, Panel A report the results with data use per category as the dependent variable. The small changes in some categories illustrate why we report these results in MB/month rather than GB/month. Panel B of Table 2-9 Regression results for the “extreme” SES and geographic groups by category, using Data as the dependent variable.is analogous to Panel A but estimated with the log of data use per category as the dependent variable. Table 2-10 is analogous to Table 2-9 but is based on the number of sessions instead of the amount of data consumed. Figure 2-7 and Figure 2-8 show the results graphically for Panel B in Table 2-9 and Table 2-10, using a bar chart similar to Figure 2-6. These figures illustrate that each geosocial group increases their usage across all categories after adopting the unlimited plan, although some groups’ increases are larger than others’.

Table 2-8 Regression results for the “extreme” SES and geographic groups, using only the category-level data

	DV: *Data*$_{it}$			Log(1+*Data*$_{it}$)		
Model	(1)	(2) Urban	(3) Rural	(4)	(5) Urban	(6) Rural
Unlimited$_{it}$	4,923.78***	4,923.78***	6,940.23***	0.50***	0.50***	0.59***
	(48.25)	(48.25)	(292.26)	(0.01)	(0.01)	(0.03)
Unlimited$_{it}$ * Urban-Low SES	2,033.89***	2,033.89***		0.21***	0.21***	
	(247.05)	(247.03)		(0.03)	(0.03)	
Unlimited$_{it}$ * Rural-High SES	2,016.44***			0.09***		
	(295.94)			(0.03)		
Unlimited$_{it}$ * Rural-Low SES	1,368.65***		-647.79	0.21***		0.12***
	(289.72)		(408.88)	(0.03)		(0.04)
Constant	8,434.49***	8,379.71***	9,160.78***	8.47***	8.47***	8.47***
	(24.64)	(25.08)	(113.64)	(0.00)	(0.00)	(0.01)
Observations	1,224,113	1,138,269	85,844	1,224,113	1,138,269	85,844
R-squared	0.26	0.26	0.24	0.33	0.33	0.34
Number of households	111,283	103,479	7,804	111,283	103,479	7,804

Notes: Fixed effects for households and months (interacted with cohort and geosocial group) included. For models 2 and 5, the baseline reference group is Urban-High SES. For models 3 and 6, the baseline reference group is Rural-High SES. Clustered standard errors (by household) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 2-9 Regression results for the "extreme" SES and geographic groups by category, using Data as the dependent variable.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Pooled	Media	Social	Tech	Sports	Business	Shopping	News	Education	Careers
Panel A – DV: $Data_{it}$										
Unlimited_{it}	4,923.78***	2,674.07***	1,000.31***	786.40***	62.08***	361.00***	16.88***	13.45***	9.10***	0.51***
	(48.25)	(34.46)	(12.63)	(12.86)	(3.20)	(5.50)	(0.64)	(0.86)	(0.87)	(0.06)
Unlimited_{it} * an-Low SES	2,033.89***	1,326.59***	459.48***	219.71***	8.80	21.64	0.64	-5.11	1.92	0.21
	(247.05)	(184.76)	(60.77)	(56.27)	(20.55)	(23.63)	(2.08)	(3.42)	(3.23)	(0.14)
Unlimited_{it} * al-High SES	2,016.44***	1,556.89***	226.55***	167.81**	-14.21	63.59**	10.15**	-0.98	6.51	0.14
	(295.94)	(232.01)	(65.41)	(70.49)	(11.73)	(28.85)	(4.93)	(4.63)	(4.04)	(0.46)
Unlimited_{it} al-Low SES	1,368.65***	1,472.19***	79.14	-86.21	-45.00***	-44.14*	-6.19**	-0.02	-1.17	0.05
	(289.72)	(229.52)	(61.44)	(64.89)	(10.65)	(23.65)	(2.81)	(3.37)	(10.52)	(0.32)
Constant	8,434.49***	3,060.85***	2,385.72***	2,000.74***	125.14***	692.04***	70.58***	72.85***	24.70***	1.86***
	(24.64)	(17.51)	(6.60)	(8.31)	(2.09)	(2.73)	(0.56)	(0.63)	(0.58)	(0.04)
R-squared	0.26	0.14	0.28	0.08	0.01	0.16	0.03	0.01	0.00	0.00
Panel B – DV: log(1+$Data_{it}$)										
Unlimited_{it}	0.50***	0.69***	0.35***	0.37***	0.29***	0.42***	0.28***	0.27***	0.24***	0.08***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
Unlimited_{it} * an-Low SES	0.21***	0.26***	0.13***	0.20***	0.03	0.13***	0.10***	0.09***	0.02	0.03**
	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Unlimited_{it} * al-High SES	0.09***	0.25***	-0.06*	0.04	-0.00	0.07***	0.07***	0.04	0.01	0.00
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Unlimited_{it} * al-Low SES	0.21***	0.38***	0.06	0.14***	-0.03	0.12***	0.08***	0.05*	0.02	0.00
	(0.03)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Constant	8.47***	6.83***	6.90***	7.07***	2.74***	5.69***	3.37***	3.03***	2.00***	0.46***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R-squared	0.33	0.33	0.42	0.25	0.07	0.35	0.18	0.08	0.12	0.03

Notes: n = 1,224,113 observations (111,283 households x 11 months). Fixed effects for households and months (interacted with cohort and geosocial group) included. The baseline reference group is Urban-High SES. Clustered standard errors (by household) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 2-10 Regression results for the "extreme" SES and geographic groups by category, using Sessions as the dependent variable.

	(1) Pooled	(2) Media	(3) Social	(4) Tech	(5) Sports	(6) Business	(7) Shopping	(8) News	(9) Education	(10) Careers
					Panel A – DV: *$Sessions_{it}$*					
$Unlimited_{it}$	5,850.78***	576.34***	802.37***	2,488.40***	81.57***	1,767.56***	71.28***	44.62***	15.45***	3.19***
	(69.07)	(6.45)	(11.73)	(32.25)	(2.71)	(23.55)	(1.93)	(1.57)	(1.37)	(0.36)
$Unlimited_{it}$ * Urban Low-SES	2,104.34***	316.61***	519.62***	758.56***	-26.36***	505.23***	31.60***	-7.35	4.38	2.05**
	(303.51)	(33.13)	(58.40)	(129.48)	(9.56)	(109.24)	(9.02)	(5.55)	(4.46)	(0.81)
$Unlimited_{it}$ * Rural High-SES	1,284.59***	321.74***	249.19***	304.25*	-18.63	395.15***	21.75	10.05	1.90	-0.83
	(392.31)	(41.85)	(71.02)	(166.22)	(17.64)	(145.54)	(13.29)	(10.09)	(7.15)	(1.06)
$Unlimited_{it}$ * Rural Low-SES	1,236.20***	263.26***	464.29***	282.87**	-53.70***	244.83*	25.82*	-0.90	9.92	-0.19
	(347.62)	(38.99)	(73.70)	(142.10)	(10.39)	(128.16)	(13.44)	(11.79)	(6.57)	(0.94)
Constant	25,430.92***	1,519.39***	3,227.49***	11,975.10***	438.37***	7,468.59***	393.24***	272.20***	120.82***	15.73***
	(41.42)	(3.66)	(6.95)	(19.15)	(1.89)	(14.86)	(1.29)	(1.04)	(0.91)	(0.22)
R-squared	0.46	0.22	0.44	0.52	0.04	0.30	0.13	0.04	0.04	0.01
					Panel B – DV: log(1+*$Sessions_{it}$*)					
$Unlimited_{it}$	0.28***	0.38***	0.31***	0.25***	0.26***	0.30***	0.25***	0.25***	0.25***	0.19***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
$Unlimited_{it}$ * Urban Low-SES	0.12***	0.19***	0.12***	0.11***	0.10***	0.14***	0.13***	0.12***	0.07***	0.07***
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
$Unlimited_{it}$ * Rural High-SES	0.04	0.12***	-0.02	0.03	0.03	0.03	0.05*	0.04	0.01	0.02
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
$Unlimited_{it}$ * Rural Low-SES	0.16***	0.27***	0.13***	0.14***	0.10***	0.18***	0.15***	0.09***	0.07**	0.01
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Constant	9.82***	6.80***	7.60***	9.11***	4.45***	8.48***	5.29***	4.66***	3.50***	1.50***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R-squared	0.33	0.24	0.50	0.42	0.10	0.25	0.22	0.08	0.19	0.06

Notes: n = 1,224,113 observations (111,283 households x 11 months). Fixed effects for households and months (interacted with cohort and geosocial group) included. The baseline reference group is Urban-High SES. Clustered standard errors (by household) in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.

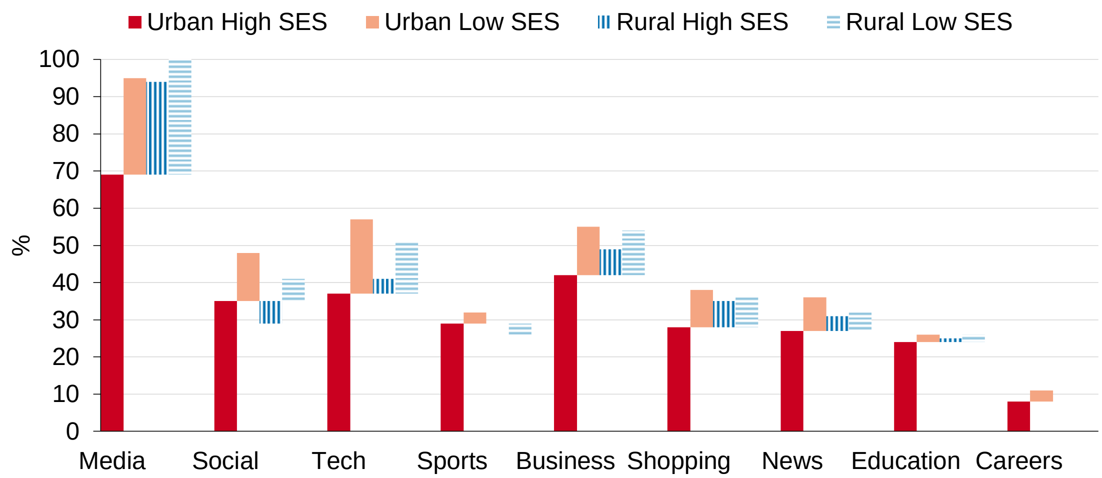

Figure 2-7 Average % increase in data use by category after adopting an unlimited plan.

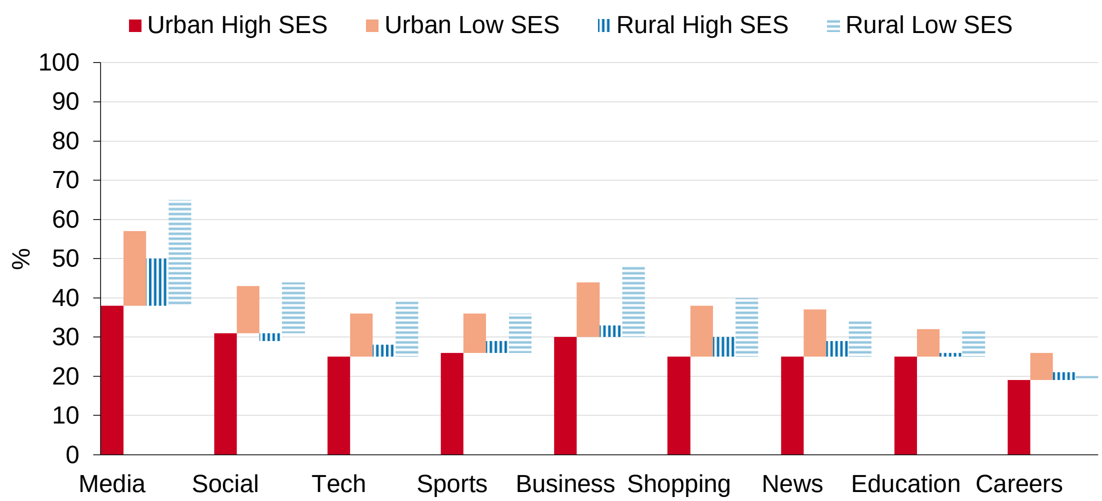

Figure 2-8 Average % increase in sessions by category after adopting an unlimited plan.

In terms of the raw increase in data consumption, Panel A of Table 2-9 shows that the Media, Social, and Technology categories account for most of the increase. They also account for most of the raw increase in sessions (see Panel A of Table 2-10), although there is also a large increase in Business sessions, which likely reflects an increased number of advertisements (see the category descriptions in Table 2-3). In terms of the percentage increases, Figure 2-7 and Figure 2-8 show that low SES households in both urban and rural areas have larger percentage increases in both data consumption and sessions (in general) compared to their high SES counterparts.

We were particularly interested in the News, Education, and Careers categories, as these may reflect socially beneficial content, at least more so than the other categories. When the raw values of $Data_{it}$ and $Sessions_{it}$ are used as the dependent variables, each geosocial group increased their use of these categories by similar amounts (in general). When the logged dependent variables are used (as illustrated in Figure 2-7 and Figure 2-8), we see significantly larger percentage increases from low SES households. For example, Urban-Low SES households had significantly larger percentage increases in sessions for the News, Education, and Careers categories compared to Urban-High SES households: 37%, 32%, and 26% vs. 25%, 25%, and 19%, respectively (see Figure 2-8). Rural-Low SES households had similar (and significant) increases in sessions for the News and Education categories: 34% and 32% vs. 25% and 25%, respectively. Urban-Low SES households also had significantly larger percentage increases in data consumed in the News and Careers categories, while Rural-Low SES households had a significantly larger percentage increase in data consumed in the News category (see Figure 2-7). However, these represent percentage increases from a fairly low baseline, such that it is not clear how meaningful they are. Also, we cannot infer from our results whether adoption of the unlimited data plans reduces the gap posited by the knowledge gap hypothesis. Although we find that Low SES households increase their consumption relative to their High SES counterparts in important content categories, which suggests that the knowledge gap might be narrowing, we have no information on how effectively households assimilate the new information that they access.

2.5 Conclusion

Can the mobile internet help close the digital divide? We studied this question by leveraging the phased adoption of unlimited mobile data plans to identify their effect on the amount and type of data consumed. Using a difference-in-differences approach coupled with coarsened exact matching, we found that adoption of unlimited plans increases data consumption for all adopters. The increase was particularly large for households on the "wrong" side of the digital divide, i.e., those with low socioeconomic status and those who live in rural areas. We believe that this is because these households lack good alternatives for high-quality internet access, due to the unavailability or high cost of fixed broadband internet service. This indicates that unlimited mobile data plans are an effective way to reduce the digital divide by increasing internet access and data consumption for traditionally underserved households. Although most of the increase was for media and entertainment content, adopting households also increased their consumption of content more likely to be socially beneficial, such as news, education, and career-related content. Although increases in these categories might not be as large as some advocates of closing the digital divide would hope, they are increases nonetheless and may help reduce inequality across socioeconomic groups and urban/rural areas over time.

The digital divide continues to be a significant public policy issue, because individuals and households without high-quality internet access may be unable to participate fully in our increasingly digital society. One way to close the digital divide is to leverage mobile technology, which may be a cost-effective way to provide high-quality internet access to households that are poorly served by fixed internet infrastructure. We show that mobile technology is indeed effective at closing the digital divide, at least when coupled with unrestricted data plans from telecommunications providers. As a result, policy makers should encourage these plans. However, if mobile broadband is to be a key ingredient in closing the digital divide, then other changes are likely necessary. For example, the small screens of mobile phones may make it difficult to access certain types of content, including educational and career-related content and services. Thus, policy makers can work with content providers to help them make their services fully accessible via mobile phone. Also, because much of the increased data used by adopters of the unlimited plans appears to be for non-enhancing pursuits, policy makers should continue to invest in educational programs about use of the internet for life enhancing purposes.

Our paper has limitations that present opportunities for future research. Although we are able to analyze the types of content that adopting households access, our content categories are relatively coarse and not necessarily mutually exclusive. For example, there is anecdotal evidence that many people use social media as a source for news (Shearer and Gottfried 2017). Future research can extend our work by using more granular measurement of the data consumed by households. This can deepen our understanding of whether, why, and to what extent increased internet access is welfare enhancing. Another limitation of our study is that we do not observe whether improved internet access leads to beneficial outcomes for the adopters.

2.6 References

Agarwal R, Animesh A, Prasad K (2009) Research note—Social interactions and the "Digital Divide": Explaining variations in internet use. *Information Systems Research* 20(2):277–294.

Angrist JD, Pischke JS (2008) *Mostly harmless econometrics: An empiricist's companion* (Princeton University Press, Princeton, NJ).

Atasoy H (2013) The effects of broadband internet expansion on labor market outcomes. *ILR Review* 66(2):315–345.

AT&T (2015) AT&T completes acquisition of DIRECTV. (July 24) http://about.att.com/story/att_completes_acquisition_of_directv.html.

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119(1):249–275.

Bonfadelli H (2002) The Internet and knowledge gaps: A theoretical and empirical investigation. *European Journal of Communication* 17(1):65–84.

Brodkin J (2016) Google Fiber division cuts staff by 9%, "pauses" fiber plans in 11 cities. *Ars Technica* (October 25) https://arstechnica.com/information-technology/2016/10/google-fiber-laying-off-9-of-staff-will-pause-plans-for-10-cities/.

Burtch G, Carnahan S, Greenwood BN (2018) Can you gig it? An empirical examination of the gig-economy and entrepreneurial activity. *Management Science* Forthcoming.

Caumont A (2013) Who's not online? 5 factors tied to the digital divide. *Pew Research Center* (November 8) http://www.pewresearch.org/fact-tank/2013/11/08/whos-not-online-5-factors-tied-to-the-digital-divide/.

Christensen G, Steinmetz A, Alcorn B, Bennett A, Woods D, Emanuel EJ (2013) The MOOC phenomenon: Who takes massive open online courses and why? *SSRN* (November 6) https://papers.ssrn.com/abstract=2350964.

Compaine B (2001) *The Digital Divide: Facing a crisis or creating a myth?* (MIT Press, Cambridge, MA).

Council of Economic Advisors (2015) Mapping the Digital Divide. (July) https://obamawhitehouse.archives.gov/sites/default/files/wh_digital_divide_issue_brief.pdf.

DiMaggio P, Hargittai E, Celeste C, Shafer S (2004) From unequal access to differentiated use: A literature review and agenda for research on digital inequality. Neckerman K, ed. *Social Inequality*. (Russell Sage Foundation, New York).

Dutton WH, Sweet PL, Rogers EM (1989) Socioeconomic status and the early diffusion of personal computing in the United States. *Social Science Computer Review* 7(3):259–271.

F.C.C. (2017a) FCC Adopts Mobility Fund Order. *Federal Communications Commission* (February 23) https://www.fcc.gov/document/fcc-adopts-mobility-fund-order.

F.C.C. (2017b) Bridging the Digital Divide for all Americans. *Federal Communications Commission* (July 27) https://www.fcc.gov/about-fcc/fcc-initiatives/bridging-digital-divide-all-americans.

Ghose A, Goldfarb A, Han SP (2012) How is the mobile internet different? Search costs and local activities. *Information Systems Research* 24(3):613–631.

Goldfarb A, Prince J (2008) Internet adoption and usage patterns are different: Implications for the digital divide. *Information Economics and Policy* 20(1):2–15.

Greenstein S (2005) The economic geography of Internet infrastructure in the United States. Cave M, Majumdar S, Vogelsang I, eds. *Handbook of telecommunication economics, Volume 2: Technology evolution and the internet*. (Elsevier, Amsterdam), 287–372.

Greenstein S (2015) *How the Internet Became Commercial: Innovation, Privatization, and the Birth of a New Network* (Princeton University Press, Princeton).

Greenstein S, Prince J (2009) Internet diffusion and the geography of the digital divide in the United States. Avgerou C, Mansell R, Quah D, Silverstone R, eds. *The Oxford Handbook of Information and Communication Technologies*. (Oxford University Press, Oxford).

Greenwood BN, Agarwal R (2015) Matching platforms and HIV incidence: An empirical nvestigation of race, gender, and socioeconomic status. *Management Science* 62(8):2281–2303.

Greenwood BN, Wattal S (2017) Show me the way to go home: An empirical investigation of ride-sharing and alcohol related motor vehicle fatalities. *MIS Quarterly* 41(1):163–188.

Hargittai E, Hinnant A (2008) Digital inequality: Differences in young adults' use of the internet. *Communication Research* 35(5):602–621.

Helsper EJ, van Deursen AJAM (2015) The third-level Digital Divide: Who benefits most from being online? Robinson L, Cotten SR, Schulz J, Hale TM, Williams A, eds.

Communication and Information Technologies Annual. Studies in Media and Communications. (Emerald Group Publishing Limited), 29–52.

Hindman DB (2000) The rural-urban Digital Divide. *Journalism & Mass Communication Quarterly* 77(3):549–560.

Hitt L, Tambe P (2007) Broadband adoption and content consumption. *Information Economics and Policy* 19(3):362–378.

Hsieh JJPA, Rai A, Keil M (2008) Understanding digital inequality: Comparing continued use behavioral models of the socio-economically advantaged and disadvantaged. *MIS Quarterly* 32(1):97–126.

Hsieh JJPA, Rai A, Keil M (2010) Addressing digital inequality for the socioeconomically disadvantaged through government initiatives: Forms of capital that affect ICT utilization. *Information Systems Research* 22(2):233–253.

Iacus SM, King G, Porro G (2012) Causal inference without balance checking: Coarsened Exact Matching. *Political Analysis* 20(1):1–24.

Kang C (2016) Bridging a Digital Divide that leaves schoolchildren behind. *The New York Times* (February 22) https://www.nytimes.com/2016/02/23/technology/fcc-internet-access-school.html.

Kolko J (2010) How broadband changes online and offline behaviors. *Information Economics and Policy* 22(2):144–152.

Kolko J (2012) Broadband and local growth. *Journal of Urban Economics* 71(1):100–113.

Levitz J, Bauerlein V (2017) Rural America is stranded in the dial-up age. *Wall Street Journal* (June 15) https://www.wsj.com/articles/rural-america-is-stranded-in-the-dial-up-age-1497535841.

Meg J (2016) Charter completes purchase of Time Warner Cable, Bright House. *Los Angeles Times* (May 18) http://www.latimes.com/entertainment/envelope/cotown/la-et-ct-charter-time-warner-cable-20160517-snap-story.html.

Monica A, Andrew P (2016) 13% of Americans don't use the internet. Who are they? *Pew Research Center* (September 7) http://www.pewresearch.org/fact-tank/2016/09/07/some-americans-dont-use-the-internet-who-are-they/.

Monica A, Horrigan JB (2016) Smartphones help those without broadband get online, but don't necessarily bridge the digital divide. *Pew Research Center* (October 3) http://www.pewresearch.org/fact-tank/2016/10/03/smartphones-help-those-without-broadband-get-online-but-dont-necessarily-bridge-the-digital-divide/.

Pai A (2017) Setting the record straight on the Digital Divide. *Federal Communications Commission* (February 7) https://www.fcc.gov/news-events/blog/2017/02/07/setting-record-straight-digital-divide.

Prieger JE (2013) The broadband digital divide and the economic benefits of mobile broadband for rural areas. *Telecommunications Policy* 37(6):483–502.

Racherla P, Mandviwalla M (2013) Moving from access to use of the information infrastructure: A multilevel sociotechnical framework. *Information Systems Research* 24(3):709–730.

Ransbotham S, Fichman RG, Gopal R, Gupta A (2016) Special section introduction—Ubiquitous IT and digital vulnerabilities. *Information Systems Research* 27(4):834–847.

Ruiz RR (2015) F.C.C. chief seeks broadband plan to aid the poor. *The New York Times* (May 28) https://www.nytimes.com/2015/05/28/business/fcc-chief-seeks-broadband-plan-to-aid-the-poor.html.

Shadish WR, Cook TD, Campbell DT (2001) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* 2nd ed. (Cengage Learning, Boston, MA).

Shearer E, Gottfried J (2017) News use across social media platforms 2017. *Pew Research Center's Journalism Project* (September 7) http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017/.

Talbot D (2016a) Furiously fast, wireless Internet access at home? It's coming. *MIT Technology Review*. Retrieved (December 4, 2017), https://www.technologyreview.com/s/601442/wireless-super-fast-internet-access-is-coming-to-your-home/.

Talbot D (2016b) The unacceptable persistence of the Digital Divide. *MIT Technology Review* (December 16) https://www.technologyreview.com/s/603083/the-unacceptable-persistence-of-the-digital-divide/.

Tichenor PJ, Donohue GA, Olien CN (1970) Mass media flow and differential growth in knowledge. *The Public Opinion Quarterly* 34(2):159–170.

Venkatesh V, Sykes TA (2012) Digital Divide initiative success in developing countries: A longitudinal field study in a village in India. *Information Systems Research* 24(2):239–260.

Viard VB, Economides N (2014) The Effect of Content on Global Internet Adoption and the Global "Digital Divide." *Management Science* 61(3):665–687.

Wei KK, Teo HH, Chan HC, Tan BCY (2010) Conceptualizing and testing a social cognitive model of the Digital Divide. *Information Systems Research* 22(1):170–187.

CHAPTER 3. SOCIAL MEDIA, FLASH SALES, AND THE MAKER MOVEMENT: AN EMPIRICAL ANALYSIS

3.1 Introduction

> "In an age of custom-fabricated, do-it-yourself product design and creation, the collective potential of a million garage tinkerers and enthusiasts is about to be unleashed, driving a resurgence of American manufacturing. A generation of 'Makers' using the Web's innovation model will help drive the next big wave in the global economy, as the new technologies of digital design and rapid prototyping gives everyone the power to invent -- creating 'the long tail of things'."
>
> Chris Anderson (2014)

> "A core part of our strategy is to find the best products in the world from emerging and established designers alike, from hyper-local to international, from a designer/maker in Brooklyn to a huge retailer in Berlin. We have sold products from designers from more than 80 countries. We take particular pride in discovering new designers and shining a spotlight on their work and helping them globally scale up their businesses. At Fab we believe that the heroes of the world are the people who dare to see things differently, makers who turn their passion into ideas and objects that can touch the lives of millions."
>
> Jason Goldberg, Founder and Chief Executive Officer, Fab (2013)

Three economic forces in the 2010s have shaped the future of e-commerce. First, the growth of the maker movement, which comprises millions of people who are willing to take big risks to start their own small businesses to create and sell handmade products, has created a plentiful supply of artisanal and handmade products (Bajarin 2014; Brit 2013; Hill 2015; Stewart 2013; Stone 2015). Second, changing consumer preference (Simon 2015) has created strong demand for goods made by these artisans and craftsmen, referred to in this article as designers[1]. Revenue from maker-driven companies in 2013 is estimated at more than $1 billion (Birkner 2015). Third, Amazon has single-handedly dominated the e-commerce market (Whaba 2015), driving its competitors to adopt innovative business methods. As branded products become more of a commodity (Jordan 2013) and improved search and recommendation engines reduce buyer search costs (De, Hu, and Rahman 2010), existing retail companies and new entrants find it harder to compete with Amazon, which almost always has the lowest price, along with shipping that is fast and often free. Jeff Jordan (2013), who is a partner at Andreessen Horowitz and serves on the boards of numerous companies, including AirBnB, argues that as branded products have become commodities, making price the key differentiator, competing with Amazon involves selling differentiated products, developing private retail label products, merchandising products differently, deploying alternative distribution strategies, and leveraging unique advantages of brick-and-mortar locations. The convergence of the maker movement, consumer demand for designer products, and the competitive drive to take on Amazon led to the founding of companies such as Fab, Etsy, and Gilt, which served as platforms that allow designers to be matched with buyers (Malone, Yates, and Benjamin 1987). Many local businesses, such as Beautiful Briny Sea (Fab 2013), continue to flourish by selling their locally made products across all 50 states using these e-commerce platforms.

Although the maker movement creates a new market for selling differentiated products, it also generates new problems for buyers and sellers. First, scholars in new product development have documented that many factors influence the success or failure of new product launches, including product advantage, market potential, competitive forces in the market, and marketing strategy (Cooper 1979; Cooper and Kleinschmidt 1987). In his influential book, Cooper (2011, p. 18) explains that the odds of a successful product launch is are about one in seven: for every 7 new product ideas, about 4 enter development, 1.5 are launched, and only 1 succeeds. Thus, even though designers create hundreds of thousands of products, they seldom know which of these are winners. Second, well-known brands often have national-level advertising and promotions that help prospective customers learn about the quality and other attributes of products. However, products manufactured by local designers have little brand recognition at the national level. Thus, buyer search costs while finding matching products from among thousands of differentiated designer products are high. Bakos (1997) shows that such high buyer search costs in markets with differentiated product offerings lead to market failure. Hence, early e-commerce companies that entered this market, such as Fab, Etsy, and Gilt, have come up with innovative business models, such as flash sales (also called limited-time sales or daily deals) and social media integration (Stambor 2012) to create excitement around the products and increase the power of word of mouth (WOM) to drive down buyer search costs. Flash sales are an e-commerce business model in which the website offers products at deep discounts for a limited period, typically from a few hours to a few days. Alexis Maybank and Alexandra Wilkis Wilson (2012, p. 6), cofounders of Gilt, explain that they created a flash sale business model that transformed shopping from a slow, leisurely activity to a competitive, addictive, urgent, thrilling rush that is delivered at the same time every day.

Even though several e-commerce companies sell products using flash sales and anecdotal evidence suggests that these types of sales can lead to a significant increase in sales and in the popularity of small-scale designers (Miller 2011), surprisingly little empirical research has been conducted to quantify their economic impact. Most research to date has focused on innovation in retail business models (Grewal et al. 2011; Sorescu et al. 2011) and on group buying sites, such as Groupon (Dholakia 2010, 2011; Edelman, Jaffe, and Kominers 2016; Li and Wu 2014), that focus on promoting businesses in their local community, primarily driving offline sales of these businesses. Some studies have been done on the impact of social media (Chen, De, and Hu 2015; Chevalier and Mayzlin 2006; Forman, Ghose, and Wiesenfeld 2008; Rishika et al. 2012), but none have examined the impact of social media on flash sales of products manufactured by designers. Researchers and practitioners alike seem to know very little about how social media affects online flash sales and how these flash sales affect the sale of products on a designer's primary website. Although nearly 40% of products launched by designers fail to sell, no research is available to advise them about the predictors of successful products. In this paper, we study these largely ignored questions related to e-commerce. Also, we explore how the impact of social media is moderated by the type of goods (search vs. experience).

Research on the relationship between social media and flash sale is lacking for at least five reasons. First, sales and social media data during limited-time flash sales are hard to measure. Constructing panel data for limited-time sales involves frequent measurement of both dependent and independent variables for thousands of products. Periodically collecting such information takes a long time and requires great computing resources. Second, most previous studies on social media have not been able to match social media data with actual product sales. Instead, they use sales rank (Chen, De, and Hu 2015; Chevalier and Mayzlin 2006; Forman, Ghose, and Wiesenfeld 2008) as a proxy for actual sales. Third, studying the moderating effect of type of goods requires data on hundreds of sales covering these product combinations. Obtaining such high-quality data is extremely difficult unless the e-commerce company itself agrees to provide it. Fourth, product heterogeneity—that is, correlation between products' unobserved characteristics and sales—and reverse causality—social media can cause sales and sales can also lead to social media activities—complicate the study of the relationship. Fifth, discovering the impact of flash sales on regular sales on the designer's primary website requires the collection of data on the primary website as well. It is doubly hard to collect data on the performance of flash sales and then match these data with the sales performance on the designer's primary website. As we describe below, we have addressed these five reasons in this article.

First, we identified one of the top five flash sale websites, Flash-commerce[2], which sells products manufactured by designers using a limited-time flash sale. We then wrote a Python-based web scraper to collect sale start/end dates, aggregate sales, and social media data periodically. Second, we developed Python based web scraping application for collecting product sales data and matched them with corresponding social media data for each product. Third, as part of this study, we collected data on 24,466 products that cover both experience and search goods over a wide price range and promotion. Fourth, because we have panel data, we used a fixed-effects panel data model to handle product heterogeneity. Also, we used two independent identification techniques—an Arellano-Bond dynamic panel data estimator and instrumental variables with a fixed-effects panel model—to identify the causal impact of social media on sales. Last, we identified and matched 24 designers with their respective primary website to retrieve their product respective sales data for the duration of our study.

Understanding the effect of flash sales and social media on the sale of designer goods has important managerial implications. First, we find that social media activities are a good predictor of which products will be successful. Designers can learn whether their new product designs will be successful by launching them using flash sales and tracking the social media activities surrounding these products within the first day of launch. The mechanisms that lead to the dissemination of popular information via social media ensures that products that are well liked by many potential customers will inspire a large amount of social media activities, which serve as predictors of success for these products. Our results also demonstrate that flash sales have a positive impact by increasing daily sales at the designer's primary website. Social media are a significant factor in spreading WOM and driving product sales, though the magnitude of impact differs for each type of social media platform, moderated by the product type. Therefore, e-commerce firms should develop customized promotion, distribution, and social media integration strategies based on the product attributes and the reach provided by each social media platform.

3.2 Literature Review

Our research draws upon the existing literature on new product development, social media, and the economics of information and advertising. In the following sections, we briefly describe key findings of each stream of literature and then explain how our work not only connects these studies but also extends our knowledge of how social media and flash sales affect the sale of designer goods.

3.2.1 Success of New Products

Scholars of new product development have documented factors, such as product advantage, market potential, competitive forces in the market, and marketing strategy (Cooper 1979; Cooper and Kleinschmidt 1987), that influence the success or failure of new product launches. Cooper (2011, p. 18), in his seminal book on new product development, explains that the odds of a successful product launch are about one in seven. Kornish and

Ulrich (2010) write that even though parallel search for innovative ideas and products might lead to redundant products, empirical tests show that such redundancy is small even for narrowly defined domains. This gives hope that even though hundreds of designers are working in parallel to create new products, there is enough scope for differentiating their products in the market. Although much prior work focused on the success of new industrial products, recent scholars have taken an interest in studying the crowdfunding of new product ideas in platforms such as Kickstarter. Kuppuswamy and Bayus (2015) examine two years of data on both successful and unsuccessful funded projects to generalize that most funding happens during the first and last week, rather than the middle of the funding cycle, even after controlling for the type of projects. They also note that the average number of backers is highest at the initial stage of funding. However, not much research has been done on the performance of new products launched by designers, a gap we intend to fill in part here.

As part of the Project NewProd, an extensive investigation into what separates successful from unsuccessful new products, Cooper (1979) finds 11 major dimensions that contribute to the success of new products: product uniqueness/superiority, market knowledge and marketing proficiency, product synergy, competitive dynamics in the market, market potential, the relative price of products, marketing strategy, marketing competitiveness, newness to the firm, strength of marketing communication, and the magnitude of investment. He argues that a key to success is knowledge of customers' wants, needs, price sensitivity, and buying behavior. Designers typically do not have the resources needed to conduct market analysis to understand customer needs when designing their products. Nor do they have the knowledge to price products competitively in the marketplace. However, the social media activities surrounding products during the flash sale might be a proxy for how well these products meet or exceed customers' wants and needs. Because prospective customers who like the design and utility of the new product tend to endorse these products on social media, and popular memes and themes spread virally via social media, we argue that such social media activities are good predictors of product success. Kuppuswamy and Bayus (2015) note that the average number of backers is highest at the first stage of the funding cycle. Because backing products via crowdfunding sites such as Kickstarter is similar to endorsing products in flash sales, we further argue that the aggregate social media activities on the first day of a flash sale will be a good predictor of product success.

H1: Social media activities associated with a product are a good predictor of that product's success or failure.

3.2.2 Flash Sales

Prior research on flash sales to date has focused on innovation in retail business models and price promotions (Grewal et al. 2011; Sorescu et al. 2011), while empirical studies have been done only on group buying sites such as Groupon (Dholakia 2010, 2011; Edelman, Jaffe, and Kominers 2016; Li and Wu 2014). Groupon is a platform that promotes businesses in their local community. In contrast, the focus of our research is other e-commerce platforms that enable local designers to sell *non-information goods* at a national level primarily online. No studies to date have examined how online flash sales affect the sale of goods online. We extend the literature by identifying how flash sales promote the sale of products on the designer's primary website and can benefit from social media integration.

Blattberg, Briesch, and Fox (1995) find that temporary price reductions cause a significant short-term spike in sales, as increasing store traffic affects the sale of both complementary and competitive product categories. As prospective customers learn about the bargains at flash sales, they learn about the products on sale while also checking out other products available for sale at regular prices. While they check out other products, they may end up buying some of those products, thereby increasing sales of products that were not part of a promotion. Some empirical evidence (Dholakia 2010, 2011; Edelman, Jaffe, and Kominers 2016) suggests that online flash sales on Groupon affect offline sales at local businesses. Extending the same argument to the online environment, we argue that hosting a flash sale online will have a positive impact on sales on the designer's primary website.

H2. Online flash sales by a designer have a positive spillover effect on sales on that designer's primary website.

3.2.3 Social Media

According to a recent survey by the Pew Research Center, 65% of adults now use social networking sites such as Facebook, a nearly tenfold increase in the past decade[3]. Although the rapid proliferation of social media usage has increased the valuation of these social media companies, views on the economic impact of popular social networking sites vary. A research study that Facebook conducted with Deloitte estimates its economic value at $225 billion (Zuckerberg 2015). At the same time, the vast majority of chief marketing officers believe that social media–based marketing contributes little to a firm's bottom line (Moorman 2016). Our research draws upon the literature on WOM, social media, and e-commerce to study the relation between social media and direct economic impact or sales of products.

Bass (1969) and Rogers (2010) show that customers who have already purchased a product spread news—both positive and negative—about that product to their connections in social networks, affecting future sales of that product. In his seminal book on the diffusion of innovation, Rogers (2010) describes the mechanism by which new products or innovations are communicated about and adopted by members of a social system. He defines the innovation decision process as the mechanism by which an individual progresses from first knowledge of a product to forming an attitude about the product, to a decision to adopt or reject it, to using the new product, and to confirming the decision by providing feedback to others. More recently, Kotler and Keller (2011, p. 228) develop a similar framework—a marketing funnel—that identifies the proportion of the potential target market at each stage in the consumer decision process: awareness, consideration, preference, action, and loyalty. At the heart of the consumer decision-making process described above is how a prospective buyer obtains information about the mere availability of a product and its quality.

Because of the growth of e-commerce, much of the exchange of information between buyers and sellers moved to online platforms, giving rise to electronic WOM. Dellarocas (2003) discusses how electronic WOM differs from traditional WOM and surveys important issues related to the design, evaluation, and use of online WOM. More recently, many papers have examined the volume, valance, and variance of such online WOM on product sales (Chevalier and Mayzlin 2006; Zhu and Zhang 2010) .

The nature of WOM expressions has evolved because of the introduction of social networking sites such as Facebook, Twitter, and Pinterest, therefore scholars have shifted their focus with respect to the electronic WOM generated through these social networking sites. Aral, Dellarocas, and Godes (2013) highlight how the new social media features are transforming the manner in which we communicate, collaborate, consume, and create information. To name a few, recent papers by Rishika et al. (2012) and Kumar et al. (2016) how social media generated at the firm level in Facebook, impacts firm's performance.

Our research is closely related to interesting papers by Li and Wu (2014) and Chen, De, and Hu (2015). Li and Wu (2014) study how social media activities such as "liking" something on Facebook and tweeting affect the sale of vouchers for local businesses and services on Groupon. They use observational data collected from Groupon and employ an Arellano-Bond dynamic generalized method of moments (GMM) model to correct for endogeneity caused by the lagged dependent and independent variables. Chen, De, and Hu (2015) examine the impact of artists' broadcasting activities on music sales on MySpace, a well-known social media platform. They use a panel-vector autoregression model to explore the relationship between broadcasting promotions on social media and the sale of music. Although these articles extend our understanding of how the new social media features work, they do not explain how social media affects the flash sale of non-information goods manufactured by designers in the maker movement.

Social media play an active role in spreading information about the availability and quality of products. First, when information about product quality is imperfect, online WOM can help in shaping customers' beliefs, especially when it comes from friends on social networks. Scholars of information systems and marketing (Dellarocas 2003; Dellarocas, Zhang, and Awad 2007) have explored the economic impact online WOM and social media, noting that social media help spread awareness of products and reduce uncertainty about the quality of products. Also, as common social contacts share similar tastes (homophily) and friends are likely to know one another's preferences (tie strength), product awareness and recommendations gained through social media tend to be more relevant. Since lesser-known designers create products that are sold in flash commerce, prospective customers incur a heavy search cost to learn about these products. Social media and the electronic WOM generated through it mitigates this high search cost spreading information on the promotions and quality of products to thousands of prospective customers in electronic social networks such as Facebook and Pinterest. Hence we argue that social media activities have a positive impact on sales.

H3: The volume of social media activities associated with a product is positively related to the sale of promoted products.

This article also offers early empirical evidence support for recent theoretical work that examines how different types of producers and consumers of social media content have different power relations online (Levina and Arriaga 2014). The unique context of our study enables us to compare different platforms and types of social media generated at different levels of product hierarchy. Social media can be generated and contained within an e-commerce website, which can be viewed only by registered users of that site, such as

Faves in Flash-commerce. We call this internal social media. Alternatively, social media can also be generated within an e-commerce website but spread using external social networking sites, such as Facebook or Pinterest. We call this external social media. Social media activities can also be generated at the product level, designer or brand level, and firm level. Also, each social media platform —Facebook, Pinterest, Twitter—has different mechanisms for spreading information. Facebook is for social networking, Twitter is used for broadcasting information to anyone interested in particular topics, and Pinterest is used as a place to curate (Chocano 2012) photos and other video content about common topics of interest for anyone interested in those topics. Our article is an early effort to compare the magnitude of impacts of internal vs. external and designer-level vs. product-level social media activities as well as comparing the impact of Facebook vs. Pinterest social media activities.

3.2.4 Search vs. Experience Goods

Philip Nelson (1970, 1974), in his seminal work on the economics of information and advertising, classifies products into search and experience goods according to consumers' ability to obtain product information before purchase. He argues that although consumers conduct minimal pre-purchase information search for experience goods, they perform extensive information search for search goods. More recently, scholars of economics, marketing, and information systems have investigated how the product type influences consumers' search, consideration set, and purchase behavior (Dimoka, Hong, and Pavlou 2012; Girard, Korgaonkar, and Silverblatt 2003; Hong and Pavlou 2014; Hsieh, Chiu, and Chiang 2005; Huang, Lurie, and Mitra 2009; Klein 1998; Klein and Ford 2003; Krishnan and Hartline 2001). Some scholars (Alba et al. 1997; Klein 1998; Peterson, Balasubramanian, and Bronnenberg 1997) have argued that on e-commerce platforms, consumers can learn about products through other people's experiences, reducing the difficulty of assessing the quality of experience goods (Lynch and Ariely 2000). However, in e-commerce, product information uncertainty and higher search cost for experience goods have been shown to be a major hurdle and challenge for e-commerce managers (Girard, Korgaonkar, and Silverblatt 2003; Hong and Pavlou 2014; Weathers, Sharma, and Wood 2007). Lee and Hosanagar (2016) examine the effect of product attributes and consumer reviews on the performance of recommender systems. Li and Wu (2014) examine how the product type moderates the impact of social media on coupons from Groupon. We extend this literature by examining how the product type—experience goods vs. search goods—moderates the impact of social media on flash sales.

Social media can affect product sales by increasing awareness of products and providing information on the quality of the product through endorsements and reviews of other members of social networks. Because designer products by nature are unique, differentiated products, consumers are likely to have imperfect information about them. Therefore, both mechanisms play a positive role in promoting sales. However, information on product quality is especially difficult to find for experience goods (Nelson 1974), as consumers discover the quality of such goods only after consuming them, whereas consumers can determine the product quality of search goods by examining them prior to purchase. The information derived from social media endorsements are less important in helping consumers update their beliefs about the product quality of search goods. At the same time, social media endorsements can serve as a signal of product quality obtained through the experiences of other users of the product. There is no reason to believe that the awareness effect would be different for search and experience goods. Because both types of goods are made by designers with little national brand awareness, and they conduct no national-level advertising, these products can be considered new to the national market. Huang, Lurie, and Mitra (2009) find that although consumers spend similar amounts of time online searching for information for search and experience goods, they spend more time per page and view fewer pages for experience goods. Search attributes such as price are objective and easy to compare, but experience attributes such as ease of maintenance or usage are subjective, with a high degree of uncertainty, and difficult to evaluate. Therefore, we argue that social media plays a stronger role in helping customers reach a purchase decision for experience goods than for search goods.

H4: The impact of social media on product sales is greater for experience goods than it is for search goods.

3.3 Research Design

3.3.1 Website Description

The data for our study come from Flash-commerce, a large e-commerce platform[4]. The company was founded in 2010 as a social networking platform, but since 2011 this site has focused on selling daily design inspirations via social commerce. Early on, it had 175,000 members, and by end of 2012, it had more than 10 million registered members. The CEO of the company claimed that nearly 90% products sold on Flash-commerce are not found on any other major e-commerce site. The firm is one of a new generation of organizations that focus on selling goods made by local designers on the national market. It aims to compete with other companies by selling products that are unique and unavailable on other popular e-commerce platforms, such as Amazon. Since its inception, the site has attracted more than 7,500 designers, selling 4.3 million products—almost 1 product every 7 seconds. By the end of 2012, it was selling more than 15,000 products, 33% more than IKEA (12,000).

Flash-commerce sells products either through traditional long-term sales or limited-time (flash) sales. Although it sells its own private label–branded products and other products, it developed an exclusive partnership with designers for long-term sales and uses flash sales for selling products from hundreds of other lesser-known designers. It launched an average of about 500 new products every day. We focus on flash sales, as it gives us a unique opportunity to understand how such promotions benefit designers. The site requires users to register before making any purchases or engaging in social media activity. Users can register using an email ID or Facebook credentials, both of which are used by the firm for promotional purposes. However, after registering, the user has the option to remain anonymous or to reveal his or her identity to other users.

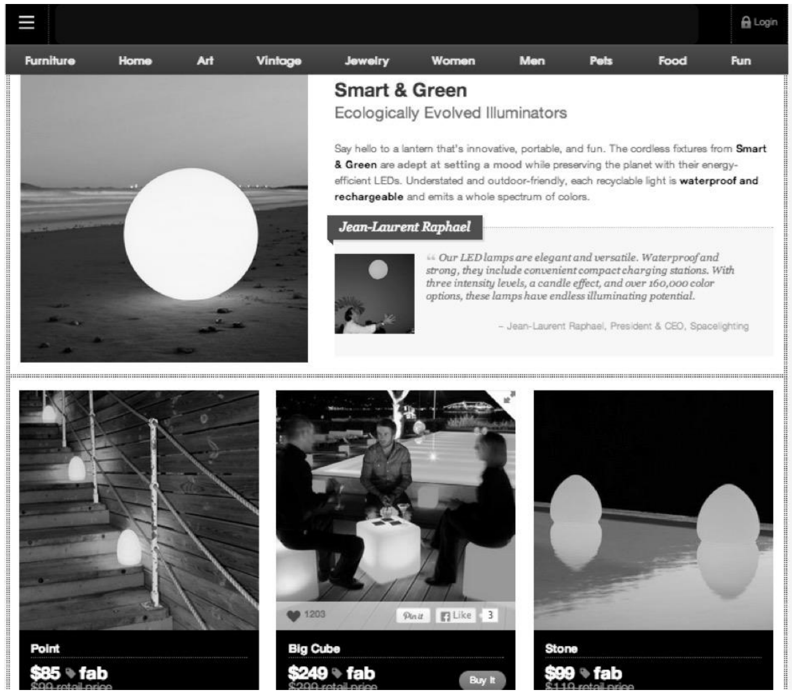

Figure 3-1 Screenshot of a typical designer page showing the first three products. The second product shows social media widgets for Faves (heart shape on the left corner) and Facebook Like button on the right-hand side showing the number of Likes so far

Figure 3-1 shows a sample webpage of a designer that sells designer light fixtures. Because such designers do not have well-known brand value, prospective customers incur heavy search costs to learn about their products. Flash-commerce mitigates this high search cost by including social media widgets at both the product and the designer level. The product called Big Cube shows three types of social media widgets: a heart-shaped button that registers Faves (favorites) on the site; a Pin-it button to post that product on Pinterest; and a Like button that registers a Like in the user's Facebook page.

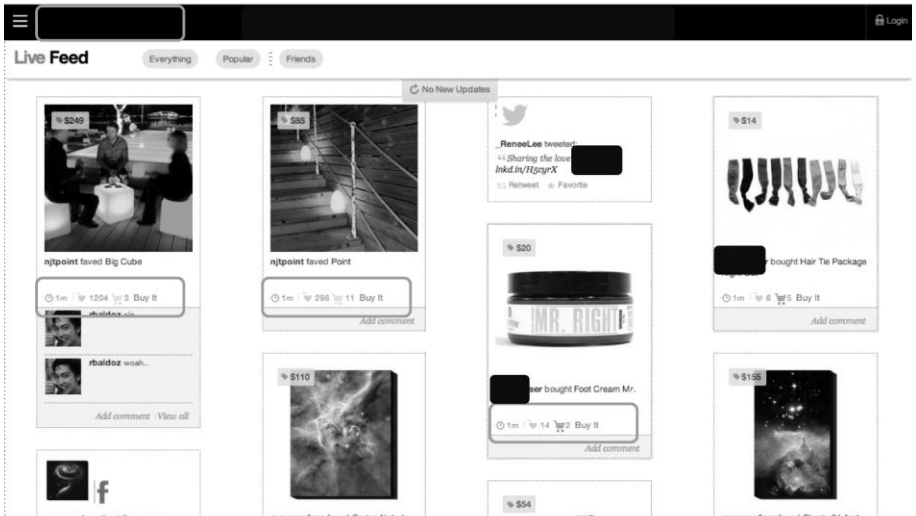

Figure 3-2 Screenshot of Live Feed page that shows social media and sales activity through an infinite scroll

In order to increase WOM, the site promotes sales and social media activities in two different ways. First, it publishes all sales and social media events on a Live Feed page, as shown in Figure 3-2, where users can see which activities are trending. This page is designed as an infinite scroll, letting users view sales and social media activities continuously. If users prefer to be anonymous, their activity will be shown as “A Flash-commerce user liked ###” or “A Flash-commerce user bought ###,” where ### is the name of the product. If users opt to make their activity public, then the same information would read as “user001 liked ###” or “user001 bought ###.” Flash-commerce also shows the cart size—that is, the cumulative quantity of products sold in these feeds. We use this cart size to track sales data for products sold on the site[5]. Second, it relies on each social media platform to broadcast its user activities on its social network. For example, when a user clicks on a Facebook Like button for a product or designer, that activity is shared on that user’s wall, as shown in Figure 3-3. Similarly, Pins are posted on the Pinterest page.

However, when a user presses heart-shaped Fave button, that activity is listed on the profile page of the Flash-commerce site, as shown in Figure 3-4.

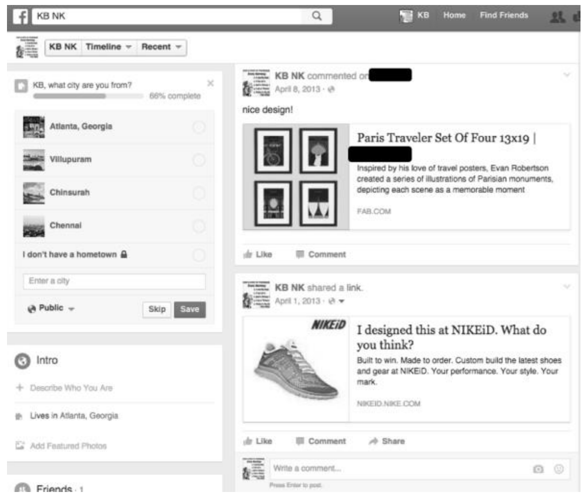

Figure 3-3 A product is shared on Facebook news feed.

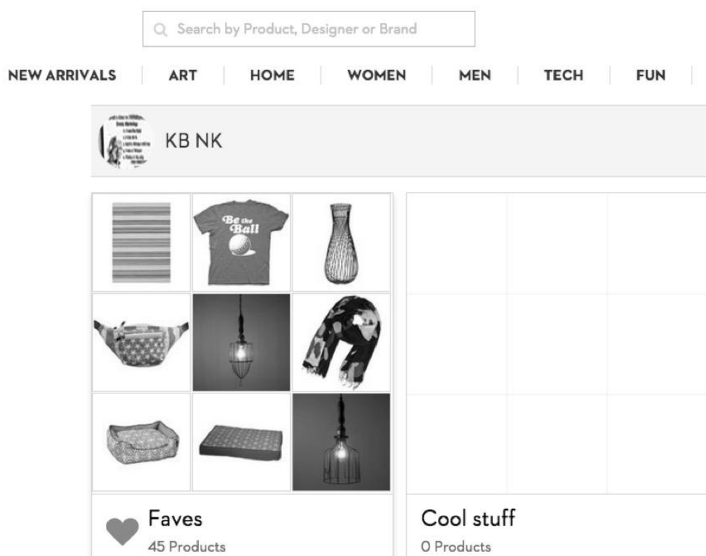

Figure 3-4 Registered users can Fave products that appear as a list on their profile page. This list can be seen by other registered users on the site

For each sale, Flash-commerce lists the name and description of the designer along with the URL of the designer's website. After analyzing all these URLs, we identify 24 designers who have a long-term presence on another e-commerce platform, Artisan-commerce[6], where they can sell their products. Interestingly, Artisan-commerce provides a sales history for all designers since their individual profile was initiated. For these 24 designers, we track daily sales on Artisan-commerce four weeks before and after these products go on sale on Flash-commerce. We normalize the launch date of the flash sale as 0 and track daily sales on Artisan-commerce 4 weeks (28 days) before and after the baseline date.

3.3.2 Data Description

We develop a Python-based web scraper to gather data on designers, products, sales quantity, and social media information from the Flash-commerce site for all products that went on flash sale from June 1, 2013, to July 31, 2013, covering more than 35,000 products. As the most popular sale period was seven days, we restricted our study to cover only products that were on flash sale for seven days. Also, due to technical issues with Internet connectivity and site maintenance, we are missing sales and social media data for a small number of products, so we excluded them from the study. At the end, we were left with 24,466 products, in categories such as books, electronics, clothes, furniture, and home goods.

Because flash sales were launched every day during the study period, we set the launch of the flash sale as the baseline and divide the sale duration into 24-hour periods for constructing our panel data. This technique is similar to the one employed by Kuppuswamy and Bayus (2015) to study cycles of crowdfunding campaigns in Kickstarter. Using our web scraper, we gather cumulative values for sales and social media variables at the end of every period. We use Stata to define this panel structure and retrieve per-period values for key analytical variables. We end up with 171,262 observations in our panel for tracking 24,466 products across 7 time periods.

Table 3-1 Descriptive statistics of key variables

	No. of Obs.	Mean	Std. Dev.	Min	Median	Max
Sales quantity$_{d,i,t}$	171,262	0.4329	1.9635	0	0	291
Product Facebook Likes$_{d,i,t}$	171,262	0.2345	2.2716	0	0	305
Designer Facebook Likes$_{d,,t}$	171,262	2.7277	33.2000	0	0	1,954
Product Faves$_{d,i,t}$	171,262	1.6556	4.2624	0	0	301
Product Pins$_{d,i,t}$	171,262	0.0444	0.2179	0	0	8
Retail price ($)	24,466	124.2174	156.2048	9	60	999
Price after discount ($)	24,466	86.9369	117.4713	5	45	999
Discount amount ($)	24,466	37.2804	63.6901	0	12	725
Discount percentage (%)	24,466	23.2511	18.4275	0	20	88
Sales quantity	24,466	3.6316	8.7463	0	1	405
Product Facebook Likes	24,466	1.9519	11.3263	0	0	690
Designer Facebook Likes	24,466	25.0199	99.0517	0	8	2,100
Product Faves	24,466	44.7691	135.6671	1	14	7,437
Product Pins	24,466	0.3876	0.6343	0	0	14
Daily sales quantity$_{d,t}$	1,368	2.4598	6.2779	0	1	64

In summary, the main time-series variables constructed for our analyses are: quantity of products sold (*Sales quantity*$_{d,i,t}$), number of Facebook Likes (*Product Facebook Likes* $_{d,i,t}$), number of Facebook Likes that are registered by the designer selling the product (*Designer Facebook Likes* $_{d,t}$), number of Faves (*Product Faves* $_{d,i,t}$), number of Pinterest Pins (*Product Pins*$_{d,i,t}$) for designer *d*, and product *i* in time period *t*. We also collect other time-invariant product attributes: the retail price of the product before the discount in $ (*Retail price*), the sale price of the product after the discount in $ (*Price after discount*), the discount percentage in % (*Discount percentage*), and the discount amount in $ (*Discount amount*). Table 3-1 lists descriptive statistics for the key variables used in this study. The average quantity of sales in each period is only 0.4329, but the maximum could reach 291 (i.e., in an extreme case, nearly 300 units of a product can be sold in a 24-hour period). The average number of product Facebook Likes is also small (0.2345), but the maximum could reach 305. Similarly, the average value of the other social media variables—designer-level Facebook Likes, product Faves, and Pinterest Pins—is relatively low, at 2.7277, 1.6556, and 0.0444 respectively. However, the maximum values can be as high as 1,954 per period. Because these five variables have very different means and standard deviations, we perform log transformation to improve the model fit. Because the minimum value of these variables is zero, we add 1 to them before the log transformation. This technique is similar to methods used in the literature (Chen, De, and Hu 2015) to mitigate the effects of highly skewed variables. Rows 10–14 show the cumulative sales values and social media variables at the end of the sale period. We notice that the average sales quantity is 3.6319 while the maximum is 690, with a median of 1. This shows that less than 1 unit is sold for 50% of the products in the sale period. The other four social media variables also follow a similarly skewed distribution.

The last row of the table gives a summary of daily sales quantity on Artisan-commerce (*Daily sales quantity$_{d,t}$*) for designer *d* in time period *t*. Since we track for 4 weeks (28 days) before and after the launch of sale on Flash-commerce, we have a total of 1,386 observations (= 24 + 28 x 2 x 24). The average daily sales quantity on Artisan-commerce is 2.4598, with a minimum of 0 and maximum of 64. This shows that a few products are successful and have high sales while 50% of products have sales of only one or zero a day.

3.4 Empirical Analysis

3.4.1 Social Media as a Predictor of Success of New Products

We employ a logistic regression model as shown in Equation (3), in which the dependent variable is the probability of final cumulative sales of a product Sales quantityd,i, which is above k. This dependent variable represents how well a product sells on Flash-commerce. In our dataset, 36.74% of the products have zero sales, 21% of the products sell just one unit, and 9.27% of the products sell two units for the duration of the sale period. So, we test for different values of k including 0, 1 (median cumulative sales), and 2 and get qualitatively similar results.

$$\text{Prob}(\text{Sales quantity}_{d,i} > k \mid \mathbf{x}) = \Lambda\,(\mathbf{x}'\,\lambda) = \Lambda\,(\lambda_0 + \lambda_1\,\text{Product Facebook Likes}_{d,i,1} + \lambda_2\,\text{Designer Facebook Likes}_{d,1} + \lambda_3\,\text{Product Faves}_{d,i,1} + \lambda_4\,\text{Product Pins}_{d,i,1} + \lambda_5\,\text{Controls}_{d,i}) \qquad (3)$$

where $\Lambda(.)$ indicates the logistic cumulative distribution function (Greene 2011).

Table 3-2 shows the results for different variants of model 1. Columns 1–3 show the results of the model with $k = 0$, and columns 4–6 show results for $k = 1$. Columns 1 and 4 list just the volume of social media activities—*Product Facebook Likes$_{d,i,1}$*, *Designer Facebook Likes$_{d,i,1}$*, *Product Faves$_{d,i,1}$*, and *Product Pins $_{d,i,1}$*—in the first 24-hour period of sale, and columns 2 and 5 have price and discount as controls: *Retail price$_{d,i}$* and *Discount percentage$_{d,i}$*. Neither the retail price nor the discount percentage changes during the sale period. In columns 3 and 6, we use the product category given by Flash-commerce to control for any association between the product type and the success of the launch.

Across all model variants, we see that all key social media variables are significant. Column 3 shows that, holding other things constant, a one-unit increase in product Facebook Likes increase the odds of success ($k > 0$) by a factor of 1.077, while a one-unit increase in product Faves and Pins increases the odds by a factor of 1.1738 and 1.2734 respectively. Interestingly, social media activities at the designer level are either insignificant (columns 1, 2, 4, 5) or have a small negative impact on success. Column 3 shows that, holding all other things constant, one-unit increase in designer Facebook Likes changes the odds by a factor of 0.9981. If we round off this factor to two digits, then it becomes 1.0, which means that the odds of the product's success remain much the same. The coefficient for designer Facebook Likes in column 6 is similar, 0.9984.

These results show that the volume of social media activities in the first 24 hours after launching a flash sale is a good predictor of success of the launch. Hence we find strong support for H1. Designers can use this feedback to gauge how well their new creations will succeed in the market, without needing to wait for a long time to gauge the market. Just launching products on a popular flash-sale platform can yield feedback within 24 hours of the launch as to whether this product will succeed in the long run. As most designers are small businesspeople without much knowledge or the resources to analyze the market potential of their product, they can utilize these flash sales as a way to test their products in a cost-effective way.

Table 3-2 Social media as a predictor of success of product launch

Dependent variable: *Prob(Sales quantity$_{d,i}$ > k /**x**)*	(1) $k = 0$	(2) $k = 0$	(3) $k = 0$	(4) $k = 1$	(5) $k = 1$	(6) $k = 1$
Product Facebook Likes $_{d,i,1}$	1.0475**	1.0709**	1.0774**	1.0384**	1.0541**	1.0596**
	(0.0234)	(0.0327)	(0.0358)	(0.0169)	(0.0227)	(0.0261)
Designer Facebook Likes $_{d,i,1}$	0.9998	0.9988	0.9981**	1.0009	0.9990	0.9984*
	(0.0007)	(0.0008)	(0.0008)	(0.0007)	(0.0007)	(0.0008)
Product Faves $_{d,i,1}$	1.1664***	1.1729***	1.1738***	1.1853***	1.1985***	1.2020***
	(0.0053)	(0.0058)	(0.0060)	(0.0051)	(0.0058)	(0.0061)
Product Pins $_{d,i,1}$	1.1181***	1.1901***	1.2734***	1.1694***	1.2915***	1.4219***
	(0.0466)	(0.0533)	(0.0578)	(0.0474)	(0.0574)	(0.0665)
Retail price$_{d,i}$ ($)		0.9934***	0.9945***		0.9876***	0.9898***
		(0.0002)	(0.0002)		(0.0004)	(0.0004)
Discount percentage$_{d,i}$ (%)		1.0103***	1.0099***		1.0128***	1.0146***
		(0.0010)	(0.0011)		(0.0013)	(0.0014)
Constant	0.8773***	1.5235***	3.6313	0.2897***	0.6499***	4.8586
	(0.0175)	(0.0482)	(4.0033)	(0.0064)	(0.0233)	(5.7843)
Observations	24,466	24,466	24,421	24,466	24,466	24,466
Control for product type	NO	NO	YES	NO	NO	YES
Pseudo-R^2	0.0990	0.1985	0.2233	0.1579	0.2906	0.3337
Regression's *p*-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Log Pseudo likelihood	-14494.765	-12894.734	-12479.433	-13992.152	-11787.782	-11071.653

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

3.4.2 Spillover Effects of Flash Sales

To find the impact of a flash sale on Flash-commerce or Artisan-commerce, we compare the average daily sales on Artisan-commerce before a flash sale is launched, with average daily sales after the sale period. If the flash sale has an impact, the average daily sales on Artisan-commerce will be statistically higher after the flash sale than before the flash sale. We estimate the following model:

$$\text{Daily sales quantity}_{d,t} = \lambda_0 + \lambda_1 \text{After flash sale indicator}_{d,t} + \mu_d + \varepsilon_{d,t} \tag{4}$$

where *After flash sale indicator*$_{d,t}$ is an indicator variable that takes a value of 1 for a designer *m* for time periods after his/her flash sale is launched; otherwise, it is assigned a value of 0.

We add designer-level fixed effects μ_d to capture the heterogeneity of each designer. If we assume that the intensity of advertising and promotion of a designer did not change during this study period, then the fixed effects would account for all the unobserved advertising and promotions done by the designer. This is not an unreasonable assumption, given that we have a short period of eight weeks during which we are looking at the spillover effects. The parameter estimate λ_1 will show us whether the flash sale increased average daily sales for the designers. However, in model 4, we do not know the duration of the impact of a flash sale. To be more precise, we assign dummy variables for each of the four weeks after a flash sale is launched: *First week after flash sale indicator*$_{d,t}$, *Second week after flash sale indicator*$_{d,t}$, *Third week after flash sale indicator*$_{d,t}$, and *Fourth week after flash sale indicator*$_{d,t}$. We also assign a dummy for the day the flash sale is launched (*Day of flash sale indicator*$_{d,t}$). Model 5 improves upon model 4 by breaking the *After flash sale indicator*$_{d,t}$ into different components with the same designer-fixed effects applied.

$$\text{Daily sales quantity}_{d,t} = \lambda_0 + \lambda_1 \text{ Day of flash sale indicator}_{d,t} + \lambda_2 \text{First week after flash sale indicator}_{d,t} + \lambda_3 \text{ Second week after flash sale indicator}_{d,t} + \lambda_4 \text{ Third week after flash sale indicator}_{d,t} + \lambda_5 \text{ Fourth week after flash sale indicator}_{d,t} + \mu_d + \varepsilon_{d,t}. \quad (5)$$

Table 3-3 Impact of flash sale on designer's regular sales

Dependent Variable: *Daily sales quantity$_{dt}$*	(1) Before vs. after flash sale	(2) Before vs. after flash sale
After flash sale indicator$_{d,t}$	0.3509**	
	(0.1385)	
Day of flash sale indicator$_{d,t}$		0.3438
		(0.9591)
First week after flash sale indicator$_{d,t}$		0.8140*
		(0.4111)
Second week after flash sale indicator$_{d,t}$		0.0521
		(0.1393)
Third week after flash sale indicator$_{d,t}$		0.3199
		(0.2021)
Fourth week after flash sale indicator$_{d,t}$		0.2188
		(0.1343)
Constant	2.2813***	2.2812***
	(0.0705)	(0.0706)
Observations	1,368	1,368
R-squared	0.0034	0.0078
Number of Designers	24	24
Designer FE	Yes	Yes

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0$

Table 3-3 presents the fixed-effects panel regression results for measuring the impact of a flash sale on sales on the primary website. It shows that flash sales have a statistically significant impact on regular sales. Column 1 gives the estimates for model 1 and indicates that launching a flash sale will increase average daily sales on the primary website by up to 0.3509 units. Column 2 shows that this impact lasts for only as long as 1 week after the flash sale, increasing average daily sales by up to 0.8140 units. The average daily sales for all designers before the start of a flash sale is 2.28125. Therefore, we find an almost one-unit increase in daily sales after launching a flash sale. Thus we find support for H2 that flash-sale promotions have a positive spillover effect on the designer's primary website. Interestingly, we find that the impact of a flash sale is felt only during the first week after the launch. It dissipates shortly after the sale period. This shows that the WOM created by a flash sale increases traffic to the designer's primary website immediately for the duration of the sale, after which the increase in traffic wears off.

3.4.3 Social Media and Product Sales

Table 3-4 Correlation between key time variant variables

		1	2	3	4	5
1	*Log(Sales quantity$_{d,i,t}$)*	1				
2	*Log(Product Facebook Likes$_{d,i,t}$)*	0.298***	1			
3	*Log(Designer Facebook Likes$_{d,t}$)*	0.310***	0.294***	1		
4	*Log(Product Faves$_{d,i,t}$)*	0.492***	0.353***	0.399***	1	
5	*Log(Product Pins$_{d,i,t}$)*	0.236***	0.210***	0.234***	0.286***	1

We first run a simple correlation between key social media variables and sales to get model-free evidence of the impact of social media on sales. Table 3-4 gives the correlation matrix for these variables. We can see that *Sales quantity$_{d,i,t}$* has a statistically significant correlation between each one of the other social media variables: *Product Facebook Likes$_{d,i,t}$* (0.298), *Designer Facebook Likes$_{d,i,t}$* (0.310), *Product Faves$_{d,i,t}$* (0.492), and *Product Pins$_{d,i,t}$* (0.236). Also, these social media variables in turn are correlated to one another. These correlations give intuitive support to H3, but we need an econometric model to separate the effect of each of these social media platforms on sales. We achieve this by running a fixed-effects panel data model, as shown in Equation (6)

$$
\begin{aligned}
&\text{Log}\big(\text{Sales quantity}_{d,i,t}\big) \\
&\quad = \sum_{s=1}^{p} \Theta_s \text{Log}(\text{Sales quanitity}_{d,i,t-s}) \\
&\quad + \sum_{s=0}^{q} \big[\lambda_s \text{Log}(\text{Product Facebook Likes}_{d,i,t-s}) \\
&\quad + \theta_s \text{Log}(\text{Designer Facebook Likes}_{d,t-s}) \\
&\quad + \sigma_s \text{Log}(\text{Product Faves}_{d,i,t-s}) \\
&\quad + \beta_s \text{Log}(\text{Product Pins}_{d,i,t-s})\big] + \mu_{d,i} + \gamma_t + \xi_{d,i,t}
\end{aligned}
\tag{6}
$$

We adapt model 6 based on the work of Chen, De, and Hu (2015) and Li and Wu (2014). Li and Wu (2014) use lagged cumulative values for social media and sales as explanatory variables, as their focus is on modeling herding behavior. Our focus is determining the impact of social media on sales. So, we use per-period sales and social media variables as explanatory variables. Chen, De, and Hu (2015) jointly model sales rank, which is a proxy for actual sales, bulletins, and friend updates as dependent variables with lagged values of these dependent variable as explanatory variables. As our focus is understanding how social media affects product sales, we just use current period sales as the dependent variable and lagged values of sales and social media as explanatory variables. Therefore model 4 uses t = 1,2…p lags of sales and t = 0,1,2..,q lags of social media variables on the right-hand side. In a seminal paper on how information spreads in Facebook, Bakshy et al. (2012, p. 523) show that the difference between the time at which subjects are first exposed to a social media link and the time at which they share is 24 hours for nearly 80% of the users. The next 24 hours increases sharing by only about 5 to 10%. This means that the vast majority of users share the information they receive via Facebook within 24 hours of viewing it. Also, as newer posts are shown first followed by older posts and there is so much information created every minute on social media every minute that older posts get buried deep in the news feed. This reduces the probability of a user acting on a post that is older than 24 hours. As Flash-commerce relies heavily on social media to spread WOM about products, we theorize that contemporaneous social media activities will have the highest effect on sales. Therefore, we simplify model 6 by using one lag of sales (t = 1) and one lag of social media variables (t = 0, 1).

Table 3-5 presents the ordinary least squares (OLS) and fixed-effects panel estimates of the effect of social media on sales for different variants of model 4. Column 1 has just the lagged sales as explanatory variable in a simple OLS model. The coefficient for Log($Sales\ quanitity_{d,i,t-1}$) is 0.3381 and is statistically significant. This shows that a 10% increase in sales in the current period increases sales in the next period by 3.381%. Column 2 shows the results for model 4 run as OLS with just contemporaneous changes in social media as explanatory variables. Column 3 combines both sales and social media in an OLS model. In both columns 2 and 3 we see that coefficient estimates for social media and sales are positive and significant. Although OLS shows a positive correlation, estimates are biased as we do not account for product heterogeneity. To overcome this issue, we include product-level fixed effects $\mu_{d,i}$ and time-fixed effects γ_t in columns 4 and 5 respectively. Column 4 shows results for model 4 with one lag of sales and no lags of social media. Column 5 gives estimates of the full model showing product- and time-fixed effects and controlling for WOM generated by sales in the previous period. We find that coefficients for all social media variables are positive and significant, indicating a strong association between sales and social media. In terms of magnitude, we find that Pinterest Pins have the highest magnitude impact. With all else held constant, a 10% increase in the *Product* $Pins_{d,i,t}$ is associated with a 1.164% change in *Sales* $quantity_{d,i,t.}$. With all else held constant, a 10% increase in the lag value of Pins *Product* $Pins_{d,i,t-1}$ is associated with a .667% increase in *Sales* $quantity_{d,i,t..}$ Similarly, a 10% increase in *Product Facebook* $Likes_{d,i,t}$, *Designer Facebook* $Likes_{d,i,t}$, and *Product* $Faves_{d,i,t}$ is associated with an increase of 0.689%, 0.301%, and 0.782% in *Sales* $quantity_{d,i,t}$ respectively, with all else held constant. We find that Pinterest Pins have the highest impact, followed by Faves, product-level Facebook Likes, and designer-level Facebook Likes. In each case, the current-period activities have a higher-magnitude impact than previous-period social media activities. This is in line with the findings of Bakshy et al. (2012) on how people respond to social media information that is current and relevant while older information is acted upon less often. The deluge of information makes it impractical to focus on older social media content. Thus the impact of lagged social media values on current period sales is much smaller, in many cases one order of magnitude smaller than that of contemporaneous social media activities. Comparing results of column 4 and 5, we find that the estimates for contemporaneous social media activities are very close to each other. For *Product* $Pins_{d,i,t}$ it is 0.1065 and 0.1164; for *Product* $Faves_{d,i,t}$ *it is* 0.0767 and 0.0782; *Product Facebook* $Likes_{d,i,t}$ it is 0.0736 and 0.0689, *Designer Facebook* $Likes_{d,i,t}$ it is 0.0334 and 0.0301 respectively. Therefore, to keep the model parsimonious, we just retain the contemporaneous terms in social media in further robustness checks.

The results in column 5 also highlight interesting difference in magnitude between different social media platforms. *Product Pins*$_{d,i,t}$ has the highest magnitude, followed by *Product Faves*$_{d,i,t}$,*Product Facebook Likes*$_{d,i,t}$, and *Designer Facebook Likes*$_{d,i,t}$. Because these are designer products, every product tends to be unique, like a work of art. Therefore, it makes sense that the more people who curate these products in Pinterest to highlight the beauty and utility of design, the more sales it generates. While *Product Faves*$_{d,i,t}$ are generated at stores on Flash-commerce, *Product Facebook Likes*$_{d,i,t}$ Likes are generated and disseminated on the public Facebook social network. Although the former is mainly used by registered users of Flash-commerce who loyally follow its products, the latter reaches millions of potential customers who may not even know of the existence of Flash-commerce or its products. Because the magnitude of the coefficients of the two is similar (0.0782 for Fave and 0.0689 for Facebook Like), we can infer that the loyalty of old customers interacting on the website balances out the possibility of reaching millions of new customers. Also, we find that the Facebook Likes of the designers has the lowest magnitude of impact on sales. We can infer that although social media activity at the designer level adds more value to the designer as a brand, however, they do not necessarily endorse the quality or utility of individual products. Overall, results in Table 3-5 provide strong support for H3 showing that social media activities are positively associated with sales.

Table 3-5 Impact of social media on sales

Dependent variable: *Log(Sales quantity$_{i,t}$)*	(1) OLS with lagged sales	(2) OLS with social media	(3) OLS with lagged sales and social media	(4) FE with lagged sales and social media	(5) FE with lagged sales and lagged social media
Log(*Sales quantity$_{d,i,t-1}$*)	0.3381***		0.2594***	-0.0108***	-0.0320***
	(0.0037)		(0.0035)	(0.0041)	(0.0040)
Log(*Product Facebook Likes$_{d,i,t}$*)		0.1530***	0.0931***	0.0736***	0.0689***
		(0.0056)	(0.0053)	(0.0060)	(0.0059)
Log(*Product Facebook Likes$_{d,i,t-1}$*)					0.0416***
					(0.0043)
Log(*Designer Facebook Likes$_{d,i,t}$*)		0.0548***	0.0242***	0.0334***	0.0301***
		(0.0017)	(0.0018)	(0.0024)	(0.0022)
Log(*Designer Facebook Likes$_{d,i,t-1}$*)					0.0059***
					(0.0014)
Log(*Product Faves$_{d,i,t}$*)		0.2326***	0.1189***	0.0767***	0.0782***
		(0.0021)	(0.0020)	(0.0024)	(0.0024)
Log(*Product Faves$_{d,i,t-1}$*)					0.0314***
					(0.0018)
Log(*Product Pins$_{d,i,t}$*)		0.2390***	0.1169***	0.1065***	0.1164***
		(0.0120)	(0.0137)	(0.0143)	(0.0147)
Log(*Product Pins$_{d,i,t-1}$*)					0.0667***
					(0.0089)
Constant	0.0655***	0.0063***	0.0127***	0.0597***	0.0594***
	(0.0008)	(0.0011)	(0.0010)	(0.0018)	(0.0045)
Observations	146,796	171,262	146,796	146,796	146,796
R-squared	0.2007	0.2758	0.2565	0.0505	0.0511
Product FE	NO	NO	NO	YES	YES
Time FE	NO	NO	NO	YES	YES
Number of products	24,466	24,466	24,466	24,466	24,466

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

3.4.4 Robustness Checks

The baseline fixed-effects panel data specification used in model 6 suffers from endogeneity in two different ways. First, the presence of a lagged dependent variable gives rise to correlation between the lag term and unobserved heterogeneity captured by the product fixed effects, which is considered part of the error term in two-way error component models. Second, there can be reverse causality between online WOM and product sales. To address these issues, we use the the dynamic Generalized Methods of Moments (GMM) to estimate first-difference equation derived from model 6. To be more specific, we use the Arellano-Bond/Blundell-Bover two-step robust GMM estimator (Arellano and Bond 1991), which uses first difference of model 4 and treats lagged variables as instruments. We follow detailed steps included in prior literature (Chen, De, and Hu 2015) to validate the assumptions needed for GMM model and to get the right specifications. First, as our data has few time periods and many panels, we conduct the Harris-Tzavalis unit-root test to verify absence of unit roots in our panel data. We get a *p*-value less than 0.01, we reject the null hypothesis and conclude that there is no unit root in our panel. Second, to get the right specification we use the Sargan specification test. We start with lag of sales ($t = 1$) and zero lags of social media variables ($t = 0$) and get a p-value > 0.05, so we fail to reject the null hypothesis that the specified lag lengths form a correct specification and the instruments are valid. Also, as we find second order autocorrelation, we only use third lag and beyond as our instruments in our GMM estimation.

In dynamic GMM estimation technique, we exploit the construction of difference equation to treat past lagged variables as instruments. Although we have been rigorous in following literature in testing all assumptions needed to get the right specification, we explore other independent ways of addressing endogeneity so that we can compare and contrast results of two methods. Since we have observational data without any exogenous shocks, finding external instruments was rather hard. So, we exploit the fact that products sold by the same designer are rather unique but related to each other. So, we treat the social media activities of other products by the same designer as an instrument for every product's social media activities. This can be considered a valid instrument, as the social media of a given product is naturally correlated with social media activities of other products of the same designer. However, as products are unique in nature, social media activities of other products of the designer do not directly affect the sale of a given product. We use these instruments to run fixed-effects instrumental variable estimation on Equation (6) with one lag of sales (t = 1) and zero lags of social media variables (t = 0). Table 3-6 shows the results for the baseline model (column 1), the Arellano-Bond model (column 2), and the fixed-effects instrumental variables model (column 3). We can see that the parameter estimates in columns 2 and 3 are similar. First, in column 2 and 3 we find the estimates for social media to be positive and significant with the exception of Designer Facebook Likesd,i,t in column 2. In column 3, we find the Product Pinsd,i,t (0.6239) has the highest magnitude followed by Product Favesd,i,t (0.0804), Product Facebook Likesd,i,t (0.0769), and Designer Facebook Likesd,i,t (0.0144). The estimates for GMM method in column 2 have highest value for Product Favesd,i,t (0.0432) followed by Product Pinsd,i,t (0.0379), Product Facebook Likesd,i,t (0.0352) and Designer Facebook Likesd,i,t (0.0054). Although we use two independent identification technique, we notice that Product Favesd,i,t and Product Pinsd,i,t are equal in magnitude if we round of estimates to 2 decimal points. The next highest magnitude is Product Facebook Likesd,i,t while Designer Facebook Likesd,i,t is either the least in magnitude or not significant. We interpret these estimates using standard log-log model method, in which a 10% increase in explanatory variable, say Product Favesd,i,t (0.0432) is associated with 10x0.0432=4.32% increase in the dependent variable. To summarize, results in Table 3-6 show that even after controlling for potential endogeneity in two independent estimation techniques, we find support for support for H3 showing that social media activities are positively associated with sales.

Table 3-6 Robustness checks with GMM and FE-IV

Dependent variable: Log(*Sales quantity*$_{i,t}$)	(1) Panel FE with controls	(2) Arellano-Bond	(3) FE-IV
Log(*Sales quantity*$_{i,t-1}$)	-0.0108***	0.3009***	0.0598***
	(0.0041)	(0.0785)	(0.0117)
Log(*Product Facebook Likes*$_{d,i,t}$)	0.0736***	0.0352***	0.0769***
	(0.0060)	(0.0068)	(0.0107)
Log(*Designer Facebook Likes*$_{d,t}$)	0.0334***	0.0054	0.0144***
	(0.0024)	(0.0042)	(0.0018)
Log(*Product Faves*$_{d,i,t}$)	0.0767***	0.0432***	0.0804***
	(0.0024)	(0.0045)	(0.0060)
Log(*Product Pins*$_{d,i,t}$)	0.1065***	0.0379**	0.6239***
	(0.0143)	(0.0149)	(0.0566)
Constant	0.0720***	0.0961***	0.0678***
	(0.0037)	(0.0110)	(0.0038)
Observations	146,796	146,796	146,796
R-squared	0.0427		
Number of products	24,466	24,466	24,466
Product FE	YES	YES	YES
Time FE	YES	YES	YES

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

3.4.5 Search vs. Experience Goods

Because thousands of products need to be classified, we follow prior literature in identifying experience and search goods. Huang, Lurie, and Mitra (2009) use Nelson's (1970, 1974) original classification to identify three categories of experience goods—automotive parts and accessories, health and beauty products, and camera equipment—and three categories of search goods— shoes, home furniture, and garden and patio equipment. They use Nelson's classification as a starting point, as it is the most widely used classification of search and experience goods, first validated by Nelson using multiple datasets. Also, numerous studies since then have used similar classification systems for their studies. Because the goal of our study is not to classify products but to examine whether the product type, as defined by the prior literature, moderates the impact of social media, we refrain from developing our own classification system, as done in some other studies (Lee and Hosanagar 2016).

We use Flash-commerce's product classification for each designer to identify product types. For search goods, we pick men's and women's shoes, furniture, and patio and garden equipment. For experience goods, we pick electronics and instruments, health and beauty, and tools, automotive and home improvement categories as they closely match with the classification system discussed earlier.

We extend model 6 by interacting a dummy variable $\textit{Experience goods indicator}_i$ that takes a value of 1 if product d,i is an experience good with each social media variable; otherwise, it takes a value of 0. As we argued in the last section, since lagged values of social media variables have a lower impact, we drop them in model 7 to keep our model parsimonious. Also, our objective here is to just find if the association between social media and sales is moderated by product type. Introducing lagged sales term makes the model dynamic introducing biased estimates as discussed in previous section. Therefore, we build Equation (7) by ignoring the lagged sales and lagged social media terms so that we can just estimate the moderating effect after controlling for time and product fixed effects.

$$\text{Log(Sales quantity}_{d,i,t}) = \lambda_0 + \lambda_2\text{Log(Product Facebook Likes}_{d,i,t}) + \lambda_3\text{Log(Product Facebook Likes}_{d,i,t})\text{x Experience goods indicator}_i + \lambda_4\text{ Log(Designer Facebook Likes}_{d,i,t}) + \lambda_5\text{ Log(Designer Facebook Likes}_{d,i,t})\text{x Experience goods indicator}_i + \lambda_6\text{ Log(Product Faves}_{d,i,t}) + \lambda_7\text{Log(Product Faves}_{d,i,t})\text{x Experience goods indicator}_i + \lambda_8\text{Log(Product Pins }_{d,i,t}) + \lambda_9\text{ Log(Product Pins }_{d,i,t})\text{x Experience goods indicator}_i + \mu_{d,i} + \gamma_t + \varepsilon_{d,i,t}. \quad (7)$$

Table 3-7 Comparison of search and experience goods

Dependent variable: Log(*Sales quantity*$_{i,t}$)	(1) Panel FE for Search Goods	(2) Panel FE for Experience Goods	(3) Panel FE with interaction term
Log(*Product Facebook Likes*$_{d,i,t}$)	0.0424***	0.1549***	0.0424***
	(0.0158)	(0.0532)	(0.0158)
Log(*Product Facebook Likes*$_{d,i,t}$) x Experience goods Indicator			0.1125**
			(0.0554)
Log(*Designer Facebook Likes*$_{d,t}$)	0.0236***	0.0061	0.0236***
	(0.0051)	(0.0095)	(0.0051)
Log(*Designer Facebook Likes*$_{d,t}$) x Experience goods Indicator			-0.0175
			(0.0107)
Log(*Product Faves*$_{d,i,t}$)	0.0741***	0.1858***	0.0741***
	(0.0066)	(0.0156)	(0.0066)
Log(*Product Faves*$_{d,i,t}$) x Experience goods Indicator			0.1116***
			(0.0169)
Log(*Product Pins*$_{d,i,t}$)	0.0499*	0.2987***	0.0499*
	(0.0265)	(0.0785)	(0.0265)
Log(*Product Pins*$_{d,i,t}$) x Experience goods Indicator			0.2488***
			(0.0827)
Constant	0.0819***	0.0791***	0.0361***
	(0.0101)	(0.0197)	(0.0058)
Observations	13,398	2,933	16,331
R-squared	0.0949	0.2819	0.1721
Number of products	1,914	419	2,333
Product FE	YES	YES	YES
Time FE	YES	YES	YES

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3-7 presents the fixed-effects panel estimates of the moderating effect of the product type on social media on sales. Column 1 shows the results of model 5 for just search goods while column 2 shows it for experience goods. Column 3 shows the combined results for both types of products. The coefficient for *Product Pins*$_{d,i,t}$ is 0.0499 for search goods, but for experience goods, it is 0.0499 + 0.2488 = 0.2987. With all else being equal, a 10% increase *Product Pins*$_{d,i,t}$ is associated with 0.499 % increase in sales for search goods and 2.987% increase in sales for experience goods. Similarly, *Product Faves*$_{d,i,t}$ and *Product Facebook Likes*$_{d,i,t}$ both have positive and significant interaction terms. Interestingly, interaction term for *Designer Facebook Likes*$_{d,i,t}$ is not significant, indicating that social media endorsements for the designer does not necessarily indicate quality of that designer's individual products. To summarize, results in Table 3-7 show support for H4, showing that product type moderates the relationship between social media and sales.

3.5 Conclusion

The maker movement is seen by many, including President Barack Obama, as a way to rekindle American manufacturing (The White House, 2015). The Obama administration has joined hands with numerous private entities in supporting and incubating a new generation of makers. However, academic research on the economics of selling millions of differentiated products is scarce. In this study, we set up an empirical framework with a unique dataset of new products sold by an e-commerce firm to study how designers can use social media and flash sales to promote and sell their products. We find that social media are a good predictor of success of new product launches. Flash sales increase visitors to designers' primary website and increase daily sales of all products and we find strong statistical evidence of a positive social media effect on sales, with product type moderating this effect.

Our findings are useful for designers, e-commerce platform managers, and policy makers. Designers can use social media to gauge which of their designs will be successful. They can also use social media not only to increase their brand awareness but also to convert this awareness into increasing their product sales. Managers of e-commerce platforms can strategically utilize social media integration to drive sales and awareness. They can use a combination of internal and external social media to leverage their loyal customer base and a social network of potential customers to spread WOM. They can more actively promote experience goods, as these are the goods about which consumers have the most uncertainty regarding product quality. Policy makers can create incentives to encourage designers and e-commerce companies to utilize social media to minimize launches of products that are unlikely to be successful as well as to increase the sales of successful products.

We believe our paper is the first step in understanding the economics of selling designer goods through e-commerce. While we have some basic results about the impact of social media and flash sales, there are a number of open questions in this area. How do designers price their products? How do they compete with other designers? How can e-commerce platforms match sellers and prospective buyers when almost each product is unique in nature? Also, our study has a number of shortcomings including the lack of randomized experiment to make strong causal inference, and dichotomous classification of search and experience goods. We hope future research could overcome some of our limitations and extend our research to answer open questions in this area.

3.6 References

Alba, Joseph, John Lynch, Barton Weitz, Chris Janiszewski, Richard Lutz, Alan Sawyer, and Stacy Wood (1997), "Interactive Home Shopping: Consumer, Retailer, and Manufacturer Incentives to Participate in Electronic Marketplaces," Journal of Marketing, 61 (3), 38–53.

Anderson, Chris (2014), Makers: The New Industrial Revolution, New York: Crown Business.

Aral, Sinan, Chrysanthos Dellarocas, and David Godes (2013), "Introduction to the Special Issue—Social Media and Business Transformation: A Framework for Research," Information Systems Research, 24 (1), 3–13.

Arellano, Manuel and Stephen Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," The Review of Economic Studies, 58 (2), 277–97.

Bajarin, Tim (2014), "Why the Maker Movement Is Important to America's Future," Time, (accessed April 11, 2016), [available at http://time.com/104210/maker-faire-maker-movement/].

Bakos, J. Yannis (1997), "Reducing Buyer Search Costs: Implications for Electronic Marketplaces," Management Science, 43 (12), 1676–92.

Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, and Lada Adamic (2012), "The Role of Social Networks in Information Diffusion," in Proceedings of the 21st International Conference on World Wide Web, WWW '12, New York, NY, USA: ACM, 519–28.

Bass, Frank M. (1969), "A New Product Growth for Model Consumer Durables," Management Science, 15 (5), 215–27.

Birkner, Christine (2015), "Makers' Market," (accessed June 4, 2016), [available at https://www.ama.org/publications/MarketingNews/Pages/makers-market.aspx].

Blattberg, Robert C., Richard Briesch, and Edward J. Fox (1995), "How Promotions Work," Marketing Science, 14 (3_supplement), G122–32.

Brit, Morin (2013), "What Is the Maker Movement and Why Should You Care?," The Huffington Post, (accessed May 20, 2016), [available at http://www.huffingtonpost.com/brit-morin/what-is-the-maker-movemen_b_3201977.html].

Chen, Hailiang, Prabuddha De, and Yu Jeffrey Hu (2015), "IT-Enabled Broadcasting in Social Media: An Empirical Study of Artists' Activities and Music Sales," Information Systems Research, 26 (3), 513–31.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," Journal of Marketing Research, 43 (3), 345–54.

Chocano, Carina (2012), "Pinterest, Tumblr and the Trouble With 'Curation,'" The New York Times, (accessed June 8, 2016), [available at http://www.nytimes.com/2012/07/22/magazine/pinterest-tumblr-and-the-trouble-with-curation.html].

Cooper, R. G. (1979), "The Dimensions of Industrial New Product Success and Failure," Journal of Marketing, 43 (3), 93–103.

——— and E. J. Kleinschmidt (1987), "New products: What separates winners from losers?," Journal of Product Innovation Management, 4 (3), 169–84.

Cooper, Robert G. (2011), Winning at New Products: Creating Value Through Innovation, New York: Basic Books.

Dellarocas, Chrysanthos (2003), "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," Management Science, 49 (10), 1407–24.

———, Xiaoquan (Michael) Zhang, and Neveen F. Awad (2007), "Exploring the value of online product reviews in forecasting sales: The case of motion pictures," Journal of Interactive Marketing, 21 (4), 23–45.

De, Prabuddha, Yu Hu, and Mohammad S. Rahman (2010), "Technology usage and online sales: An empirical study," Management Science, 56 (11), 1930–45.

Dholakia, Utpal M. (2010), "How effective are Groupon promotions for businesses?," Available at SSRN 1696327.

——— (2011), "What makes Groupon promotions profitable for businesses?," Available at SSRN 1790414.

Dimoka, Angelika, Yili Hong, and Paul A. Pavlou (2012), "On product uncertainty in online markets: Theory and evidence," MIS quarterly, 36 (2), 395–426.

Edelman, Benjamin, Sonia Jaffe, and Scott Duke Kominers (2016), "To Groupon or not to Groupon: The profitability of deep discounts," Marketing Letters, 27 (1), 39–53.

Fab (2013), "Fab Designer Story - Beautiful Briny Sea," (accessed May 20, 2016), [available at https://www.youtube.com/watch?v=GTkacBT1CEU&feature=youtu.be].

Forman, Chris, Anindya Ghose, and Batia Wiesenfeld (2008), "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," Information Systems Research, 19 (3), 291–313.

Girard, Tulay, Pradeep Korgaonkar, and Ronnie Silverblatt (2003), "Relationship of Type of Product, Shopping Orientations, and Demographics with Preference for Shopping on the Internet," Journal of Business and Psychology, 18 (1), 101–20.

Goldberg, Jason (2013), "Letters to Employees," (accessed May 26, 2016), [available at http://fab.com/ceoletters/2013/0930/].

Greene, William H. (2011), Econometric Analysis, Boston: Prentice Hall.

Grewal, Dhruv, Kusum L. Ailawadi, Dinesh Gauri, Kevin Hall, Praveen Kopalle, and Jane R. Robertson (2011), "Innovations in Retail Pricing and Promotions," Journal of Retailing, Innovations in Retailing, 87, Supplement 1, S43–52.

Hill, Paul (2015), "Ten statistics that reveal the size and scope of the Maker Movement – eXtension," (accessed June 4, 2016), [available at https://extension.org/2015/11/03/ten-statistics-that-reveal-the-size-and-scope-of-the-maker-movement/].

Hong, Yili and Paul A. Pavlou (2014), "Product fit uncertainty in online markets: nature, effects, and antecedents," Information Systems Research, 25 (2), 328–44.

Hsieh, Yi-Ching, Hung-Chang Chiu, and Mei-Yi Chiang (2005), "Maintaining a committed online customer: A study across search-experience-credence products," Journal of Retailing, 81 (1), 75–82.

Huang, Peng, Nicholas H Lurie, and Sabyasachi Mitra (2009), "Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods," Journal of Marketing, 73 (2), 55–69.

Jordan, Jeff (2013), "How to compete with Amazon - Fortune," (accessed June 4, 2016), [available at http://fortune.com/2013/10/24/how-to-compete-with-amazon/].

Klein, Lisa R. (1998), "Evaluating the Potential of Interactive Media through a New Lens: Search versus Experience Goods," Journal of Business Research, 41 (3), 195–203.

——— and Gary T. Ford (2003), "Consumer search for information in the digital age: An empirical study of prepurchase search for automobiles," Journal of Interactive Marketing, 17 (3), 29–49.

Kornish, Laura J. and Karl T. Ulrich (2010), "Opportunity Spaces in Innovation: Empirical Analysis of Large Samples of Ideas," Management Science, 57 (1), 107–28.

Kotler, Philip and Kevin Lane Keller (2011), Marketing Management, Prentice Hall.

Krishnan, Balaji C. and Michael D. Hartline (2001), "Brand equity: is it more important in services?," Journal of Services Marketing, 15 (5), 328–42.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P. K. Kannan (2016), “From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior,” Journal of Marketing, 80 (1), 7–25.

Kuppuswamy, Venkat and Barry L. Bayus (2015), “Crowdfunding Creative Ideas: The Dynamics of Project Backers in Kickstarter,” SSRN Scholarly Paper, Rochester, NY: Social Science Research Network.

Lee, Dokyun and Kartik Hosanagar (2016), “When Do Recommender Systems Work the Best?: The Moderating Effects of Product Attributes and Consumer Reviews on Recommender Performance,” in Proceedings of the 25th International Conference on World Wide Web, WWW ’16, Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 85–97.

Levina, Natalia and Manuel Arriaga (2014), “Distinction and status production on user-generated content platforms: Using Bourdieu’s theory of cultural production to understand social dynamics in online fields,” Information Systems Research, 25 (3), 468–88.

Li, Xitong and Lynn Wu (2014), “Herding and Social Media Word-of-Mouth: Evidence from Groupon,” SSRN Scholarly Paper, Rochester, NY: Social Science Research Network.

Lynch, John G. and Dan Ariely (2000), “Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution,” Marketing Science, 19 (1), 83–103.

Malone, Thomas W., Joanne Yates, and Robert I. Benjamin (1987), “Electronic Markets and Electronic Hierarchies,” Communications of the ACM, 30 (6), 484–97.

Maybank, Alexis and Alexandra Wilkis Wilson (2012), By Invitation Only: How We Built Gilt and Changed the Way Millions Shop, Portfolio.

Miller, Claire Cain (2011), “Brands Find That Flash Sales Online Are Useful as Marketing,” The New York Times.

Moorman, Christine (2016), “The Social Media Spend-Impact Disconnect,” The CMO Survey, (accessed June 3, 2016), [available at https://cmosurvey.org/blog/the-social-media-spend-impact-disconnect/].

Nelson, Phillip (1970), “Information and Consumer Behavior,” Journal of Political Economy, 78 (2), 311–29.

——— (1974), “Advertising as Information,” Journal of Political Economy, 82 (4), 729–54.

Peterson, Robert A., Sridhar Balasubramanian, and Bart J. Bronnenberg (1997), “Exploring the implications of the Internet for consumer marketing,” Journal of the Academy of Marketing science, 25 (4), 329–46.

Rishika, Rishika, Ashish Kumar, Ramkumar Janakiraman, and Ram Bezawada (2012), "The Effect of Customers' Social Media Participation on Customer Visit Frequency and Profitability: An Empirical Investigation," Information Systems Research, 24 (1), 108–27.

Rogers, Everett M. (2010), Diffusion of Innovations, New York: Free Press.

Simon, Ruth (2015), "Big Retailers Seek Small-Artist Touch," Wall Street Journal, (accessed June 4, 2016), [available at http://www.wsj.com/articles/big-retailers-seek-small-artist-touch-1448588044].

Sorescu, Alina, Ruud T. Frambach, Jagdip Singh, Arvind Rangaswamy, and Cheryl Bridges (2011), "Innovations in Retail Business Models," Journal of Retailing, 87, Supplement 1, S3–16.

Stambor, Zak (2012), "Fab.com leads the pack in social commerce," (accessed May 20, 2016), [available at https://www.internetretailer.com/2012/12/11/fabcom-leads-pack-social-commerce].

Stewart, Martha (2013), "Martha Stewart: Meet USA's new entrepreneurs," USA TODAY, (accessed June 4, 2016), [available at http://www.usatoday.com/story/money/business/2013/10/14/martha-stewart-column-meet-the-makers/2980701/].

Stone, Zara (2015), "The Maker Movement Is Taking Over America. Here's How...," The Hustle, (accessed June 4, 2016), [available at http://thehustle.co/the-diy-maker-movement-survives-by-doing-the-opposite-of-whats-smart].

The White House, Office of the Press Secretary (2015), "FACT SHEET: New Commitments in Support of the President's Nation of Makers Initiative," whitehouse.gov, (accessed July 7, 2016), [available at https://www.whitehouse.gov/the-press-office/2015/06/12/fact-sheet-new-commitments-support-president%E2%80%99s-nation-makers-initiative].

Weathers, Danny, Subhash Sharma, and Stacy L. Wood (2007), "Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods," Journal of Retailing, 83 (4), 393–401.

Whaba, Phil (2015), "Amazon Has Huge Lead Over Walmart, Target in Online Retail," (accessed June 4, 2016), [available at http://fortune.com/2015/11/06/amazon-retailers-ecommerce/].

Zhu, Feng and Xiaoquan (Michael) Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," Journal of Marketing, 74 (2), 133–48.

Zuckerberg, Mark (2015), "Mark Zuckerberg - A new economic report found that Facebook...," (accessed June 3, 2016), [available at https://www.facebook.com/zuck/posts/10101865957208591]

.

CHAPTER 4. WI-FI VS. MOBILE INTERNET – CONSUMER PREFERENCE FOR LAST MILE DELIVERY OF INTERNET: EVIDENCE FROM A LARGE-SCALE FIELD EXPERIMENT

4.1 Introduction

> "For the 4 billion people who are not on the internet today, there are three major obstacles. The first is **availability** of networks. That is, even if they had a phone, they [can't] get a signal because there's no fiber, or there's no mobile broadband network where they are…. Then, there's **affordability** of the network. There's another billion or so people where technically there may be a signal that they can pick up, but they can't afford to use it…. And then the third pillar, which is in a lot of ways the biggest, is about a couple billion people. For them, the issue is not availability, or affordability, but **awareness**."
>
> Mark Zuckerberg, Chief Executive Officer, Facebook (Newton 2016)

Three economic forces in the 2010s have shaped the growth and future of internet. First, mobile internet usage overtakes desktop as most used digital platform (Dreyer 2015), prompting technology companies to design applications and websites specifically for mobile first. Second, as valuation of tech companies such as Facebook and Google increasingly rely on number of active users, these companies are coming up with innovative ways of providing internet to masses. Connecting millions of new users and engaging them in internet enabled markets benefit both consumers and firms by increase overall demand for physical and information goods. Google launched its division 'Google Fiber' (Bergen 2016) to provide gigabit internet through *fiber-optic cables*. Facebook is using drones to deliver internet *wirelessly* over large areas (Newton 2016). Third, both telecommunication companies and consumers now have a number of last mile connectivity options that include mobile networks, Wi-Fi hotspots, wired broadband (cable, DSL) and gigabit fiber. Ericsson (Qureshi 2015) estimates that by 2020 broadband through *mobile networks* will reach 90% of the world population. Alternatively, local businesses such as coffee shops and airports offer *Wi-Fi hotspots* often free of charge to customers by collaborating with Internet Service Providers (ISP) (Barbagallo and Kharif 2015).

In this chapter, we will focus on two last-mile internet connectivity option for mobile phones – *mobile networks and Wi-Fi networks*. Greenstein (2005) gives a brief history as well as the infrastructure that makes up internet. The **last mile**[5] refers to the portion of the telecommunications network chain that physically reaches the end-user's premises either via copper wire as is the case for DSL lines, coaxial cables for cable TV/internet lines, or cell towers linking cell phones as is the case in mobile internet. The last mile channel largely determines the cost, security, reliability, speed and the overall quality of the internet experience for the customer. Mobile networks are expensive to build and maintain but provide a secure connectivity over a large geographical area. For consumers, mobile data plans that give access to these mobile networks are often very expensive with high overage charges. On the other hand, Wi-Fi hotspots are cheaper for both companies and consumers while providing much higher speeds compared to mobile networks, although with limited range. Also, from a policy perspective, the firm administering last mile service might gain significant competitive advantage. For example, Facebook's Free Basics internet restricts access to just a few hundred web sites only (Van Boom 2016). Verizon launched Go90 app through which its customers can stream content for free (Ravenscraft 2016).

[5] https://en.wikipedia.org/wiki/Last_mile

While there is ample evidence of a race to deliver internet through different channels, surprisingly little empirical research has been conducted to explore how firms deliver internet and how internet usage in one last-mile channel impacts the usage in other channels.

There are several reasons for the lack of research in this area. It is very hard to get individual level panel data on data consumption in two last-mile channels, in part due to the challenges coordinating the two companies that provide the connectivity to track usage of same cohort of customers in both channels. Also, because customer choice of last-mile channel is endogenous, identifying the effect of change in policies is rather difficult.

We overcome these problems by directly partnering with a leading U.S. telecommunications provider[6] (referred to as the "company"). Our partner company provides both mobile internet and Wi-Fi services in more than 40,000 hotspots across the US. As per its policies, wireless subscribers who are in good standing are automatically given access to these hotspots free of cost. However, as the company does not advertise the availability of these free hotspots, there is low awareness of this benefit even among the company's own employees. We exploit this low-level awareness by designing and implementing a large-scale randomized field experiment that increases awareness of these hotspots.

We find the promoting free Wi-Fi hotspots increases both (paid) mobile data usage and (free) Wi-Fi data usage. Heavy mobile data users tend to increase Wi-Fi usage a lot more compared to others. On the other hand, Wi-Fi increase in usage is much determined

[6] The company wishes to remain anonymous

by the type of businesses (locations) that offer the hotspots. Understanding these nuances would help telecommunication companies, urban/city planners to locate Wi-Fi and mobile infrastructure to serve the needs of the society.

4.2 Literature Review

Our research draws upon the existing literature on economics of diffusion of internet, traditional wireless phone services and emerging literature on consumer usage of mobile data and smartphone apps. Researchers in IS and Economics have established that diffusion and adoption of internet has a positive economic impact (Forman et al. 2005, 2012). Xu et al. (2015)study how mobile users plan their data consumption, given a limited data limit for every billing cycle. Ghose and Han (2011) investigate the relationship between content creation and usage for users using the mobile internet. Niculescu and Whang (2012) examine the co-diffusion process of the adoption of wireless voice and mobile data services in Japan. Xu et al. (2014) explore the complementary between the introduction of a mobile app and website visits for news media. Xu et al. (2018) examine the competition between fixed-line and mobile internet. They find that speed of fixed-line internet and socio-economic factors determine the adoption and usage of mobile internet.

In contrast, our paper is the first to study consumer usage of two wireless last-mile internet channels – mobile network and Wi-Fi. Due to our field experiment design, we eliminate most of the endogeneity concerns that arises due to self-selection of internet channels and the highly competitive nature of the mobile industry. We exogenously vary the awareness on the availability of free Wi-Fi hotspots so that we can explore how this awareness impacts both the usage of internet in these Wi-Fi hotspots and mobile internet.

4.2.1 Mobile data usage

Xu et al. (2014) establish that subscribers are either myopic or forward looking, with at least some of the consumers planning their data consumption in order to maximize their overall utility. Xu et al. (2018) find that subscribers choose between mobile internet and fixed line by examining the benefits provided by each channel. Extending these arguments, given an option between mobile network and free Wi-Fi hotspots, subscribers would carefully examine the benefits and costs before adopting and using each of these services. As a cell phone can connect to one primary network at any given time, one can argue that spending a lot of time connected to Wi-Fi hotspots would decrease the mobile data usage. As Wi-Fi hotspots offer free data usage, promoting them would cannibalize usage of paid mobile data. One could also argue that as subscribers might increase mobile internet usage as they get accustomed to the increased internet usage in their mobile phones via in free Wi-Fi hotspots. Eventually, as they get trained to use this service, such an action becomes sticky while they are in mobile networks too. For example, if a subscriber is watching a movie in free hotspot and suddenly walks out of the hotspot, usage may spill over to the mobile network. Such a spillover aggregated over many time periods would create a strong positive association between free Wi-Fi hotspot usage and mobile data usage.

We argue that the second mechanism that trains and sticks active mobile internet usage from hotspot to mobile network would eventually dominate the relationship between the two channels. Hence, we argue that the promoting Wi-Fi hotspots would increase mobile data usage.

H1: Increase in awareness of availability of free Wi-Fi hotspots is positively associated with mobile data usage.

4.2.2 Wi-Fi Usage

The recent increase in availability of mobile location data has led to the new field of urban sensing that seeks to understand how people interact with physical infrastructure. Calabrese et al. (2014) provide a summary of recent research in this area and outline how mobile location data can be used to study population distribution, type of activities performed in different locations of the city, travel patterns and geographic social networks. In our context, we are interested in finding how promoting free Wi-Fi hotspots will impact usage of Wi-Fi data. A hotspot in a coffee shop will have a different pattern of usage compared to say airport or retail mall. We argue that not all places are created equal. Therefore, the impact of our promotion on Wi-Fi usage will be moderated by the type of location/business that the hotspot is located. Locations where users spend a lot of time waiting for their main events such as airports, convention centres etc., there will be a positive association between increase in awareness and Wi-Fi usage. In other location types where users want less distraction, there will be a decrease in Wi-Fi usage.

H2: The location of Wi-Fi hotspot will play a moderating role in the relationship between increase in awareness of availability of free Wi-Fi hotspots and Wi-Fi data usage.

4.3 Research Context

To explore our research questions, we partnered with a large U.S. telecommunications provider[7] (referred to as the "company") that provides a full-range of telecommunication services, including wireless voice and data service and TV service. This company provides more than 40,000 Wi-Fi hotspots throughout the US in a number of publicly accessible locations, including restaurants, airports, retail shops etc. In addition to serving the customers of the individual establishments, it also provides this service free of charge to all of its wireless customers. However, this company does not advertise or promote this free service actively to its wireless customers. One of the key reasons was that promoting free Wi-Fi hotspots might drive down usage of paid mobile data by its wireless subscribers.

To better understand how usage of Wi-Fi hotspots is related to mobile data consumption, we designed and implemented a field experiment in June 2015 in the North East market region. We worked with the company to randomly select one million customers whose billing address was located in the North East market region of the US. We selected only those customers who are not past due and who use smartphones. We allocated them randomly into two groups – a treatment and control group with 500,000 subscribers each. In order to prevent leakage of information between the groups, we retained subscribers in the same account to be in the same group. Every subscriber in the company is associated with a unique subscriber id and an account it. On June 30th 2015, subscribers in treatment group were sent a promotional text message with the following content:

[7] The company wishes to remain anonymous

ABC FREE MSG: We appreciate you - ABC offers FREE Wi-Fi nationwide. Connect and enjoy. Go to ABC.com/WiFihotspots to learn more[8].

When a subscriber clicked on the link given in the text message from the mobile phone, the mobile phone browser was re-directed to a website that described more about the company's free Wi-Fi hotspots. This included availability of these hotspots near the subscriber's location based on the geo-location provided by the mobile phone, troubleshooting guide etc. The company setup these hotspots such that whenever its wireless subscribers in good standing took their phone within the wireless hotspot, their phone will *automatically* connect to the Wi-Fi network. This was done for several reasons. First, the company wanted to offload at least some of the mobile data to the Wi-Fi networks to reduce congestion in mobile networks. Second, additional prompts were not provided to give a seamless trouble-free experience for the subscribers. However, the subscriber always has the option to turn off Wi-Fi altogether or block the free hotspot by setting it in the phone or set some other network as the preferred Wi-Fi network. In our setup, as we randomly choose treatment and control groups, we will have similar distribution of subscribers in both the groups, helping us achieve unbiased estimates. Also, in our experiment, we examine the Wi-Fi hotspots provided by our company. We do not consider Wi-Fi hotspots either provided by other companies or those available at home. Henceforth, Wi-Fi hotspots, Wi-Fi, or hotspots refer only to the free Wi-Fi hotspots provided by our company that is being considered in this study.

[8] Name of the mobile service provider has been changed to 'ABC' to maintain anonymity as per requirements of our NDA with them

4.4 Empirical setup

In order to check if our manipulation effectively reached the targeted subscribers, we worked with the company's e-commerce team to determine the number of unique clicks for the URL in the text message. We found that between July 1st week and end of August 2015, about 25% of the subscribers in treated group had clicked on the link and explored the promotional offer. We were unable to get the total number of treated subscribers that read the text message as we did not have access to that data. As only a fraction of all users who viewed the text message would have clicked on the URL in the message, we have enough evidence that our manipulation was effective.

Table 4-1Summary statistics for accounts in July 2015

Descriptive statistics - June 2015					
Variable	Mean	Min	Median	Max	St. Dev.
Mobile data (MB/Month)	######	0	######	########	######
Wi-Fi data (MB/Month)	37.594	0	0.208	74,562.85	335.39
Wi-Fi duration (Hour/Month)	3.124	0	0.65	656.71	10.743
Wi-Fi venues (count/Month)	1.846	0	1	67	2.497
Wi-Fi sessions (Count/Month)	4.365	0	1	1,888	12.658
Rural Population (%/Zip)	0.162	0	0.008	1	0.282
Observations	891,342 accounts				

We work with our partner company to collect rich data about these subscribers aggregated at their accounts. Table 4-1 shows the summary statistics for these accounts aggregated for the month of July 2015 for a total of 891,342 accounts. *Mobile data* gives the monthly consumption of data in MB in the mobile network that is included as part of the paid data plan. On average, each account consumes about 3,117.18 MB of data per month, while there is a large heterogeneity in data consumption with minimum of zero and maximum of 286,698.9 MB. The next four rows give measure the data consumption in Wi-Fi hotspots. *Wi-Fi duration* gives the measure of hotspot sessions in hours aggregated per month. Note that as the subscriber's phone automatically connects to the Wi-Fi hotspots, the duration of Wi-Fi session at a hotspot gives us a good proxy for the time spent by the user at that Wi-Fi hotspot location. On average, subscribers in the accounts spent about 3.12 hour per month at hotspot locations, while the usage is highly skewed. Some subscribers spend zero minutes per month while some others spend as much as 700 hours per month. *Wi-Fi data gives* a measure of total data in MB utilized by each accout at the hotspot locations aggregated for each month. On average, subscribers in the account use around 38 MB/month, which is about 1 % of the mean mobile data consumption. Similar to mobile data usage, Wi-Fi data usage is also skewed. *Wi-Fi venues* measures the unique count of hotspots visited by a subscriber every month. Each hotspot location is associated with a unique identifier. We count the distinct values of this identifier to compute this variable. On average, subscribers visit around 2 unique hotspot locations every month, while some don't visit any. There are some subscribers that visit as many as 59 hotspot locations every month showing heterogeneity among our user base. *Wi-Fi sessions* measures the number of Wi-Fi sessions started by subscribers in these accounts every month. On average, there are around 4.3 sessions every month, with minimum being zero and maximum at 1888. *Rural population* gives a measure of fraction of rural population in the billing zip code of each subscriber. We get this data from the Census database to determine if the subscriber comes from a urban or rural area. Note that we have one billing zip code for each account. It is possible that some subscribers don't live at the billing zip code. However, as the distribution of such users is bound to be same in both control and treated group, we will be able to control for this bias.

While running any experiment, one of the primary concern is whether we randomly assigned users in treatment in control group. We check this by obtaining Wi-Fi and mobile usage for the baseline month of June 2015 and regressing these variables against treatment assignment. If we indeed performed random assignment, then none of the explanatory variables should be significant impact on the treatment assignment. Table 4-2 shows the results of this regression. Column 1 was run with just the variables described in the summary statistics in prior section. In column 2 we added more demographic variables obtained for each account. In both cases, we find none of these variables have significant predictive power on the treatment assignment. Therefore, we conclude that our treatment assignment is indeed random.

Also, since the mobile and Wi-Fi usage as well as time spent in hotspots are skewed, we take natural log transformations. Because the minimum value of these variables is zero, we add 1 to them before the log transformation. This technique is similar to methods used in the literature (Chen et al. 2015) to mitigate the effects of highly skewed variables.

Table 4-2 Verification of random assignment

Dependent variable	Treatment Indicator	
	(1)	(2)
Mobile data	0.000***	0.000***
	(0.000)	(0.000)
Wi-Fi data	-0.000	-0.000
	(0.000)	(0.000)
Wi-Fi duration	0.000	0.000
	(0.000)	(0.000)
Wi-Fi venues	0.001***	0.001***
	(0.000)	(0.000)
Wi-Fi sessions	0.000	0.000
	(0.000)	(0.000)
Constant	0.490***	0.497***
	(0.001)	(0.002)
Observations	891,106	890,212
R^2	0.0001	0.0002
Additional Controls	NO	YES
F Statistic	14.899***	20.141***
Note:	*p**p***p<0.01	

Since we have the advantage of reducing endogeneity by performing an experiment, we use a simple cross-sectional regression model to estimate the impact of increase in awareness on mobile and Wi-Fi data usage as shown in equation 8.

$$\text{Usage}_i = \lambda_0 + \lambda_1 \text{ Treatment Indicator}_i + \lambda_2 \text{ Controls}_i + \varepsilon_i. \quad (8)$$

The subscript *i* represents an account as the unit of observation for the month of July 2015. The dependent variable $Usage_i$ measures the data used in the mobile phone. We run this model with either mobile or Wi-Fi data usage as the dependent variable. The *Treatment Indicator*$_i$ is a dummy variable that takes a value of 1 for accounts in treated group and 0 otherwise. Our primary objective is to estimate the value of λ_1 that determines the impact of our promotion on outcomes under study. We also add many controls such as number of lines per account, monthly recurring charge, distinct Wi-Fi venues etc. for additional robustness checks.

4.5 Results

4.5.1 Mobile usage

Table 4-3 shows the results for estimates for equation 8 with mobile data as the dependent variable in the first two columns and log transformed mobile data usage in columns 3 and 4. While column 1 and 3 are run without any additional controls, columns 2 and 4 are run with additional controls to increase the explanatory power (R-Square) of the model. In all the four estimates, we find that there is a significant and positive impact of increase in awareness on mobile data usage. In terms of raw numbers, each account sees an average increase of about 75 MB/month due to treatment. Although this is small in magnitude compared to total mobile data usage (about 3000 MB/Month), we should note that we only account for a small fraction of total Wi-Fi hotspots available to the subscribers. While this is indeed a limitation of our study, one could argue that accounting for all other hotspots would have a higher magnitude impact on mobile data usage.

As an additional robustness check, we ran a quantile regression to verify if the impact of treatment varies based on the level of mobile data usage. We find evidence that higher mobile data users seem to have a higher impact of treatment (about 200 MB) when compared to low mobile data users.

Table 4-3 Impact of Wi-Fi promotion of mobile data usage

	Dependent variable:			
	Mobile data (MB/Month)		Log(1+Mobile data) (MB/Month)	
	(1)	(2)	(3)	(4)
Treatment Indicator	74.476^{***}	75.079^{***}	0.023^{***}	0.019^{***}
	(10.143)	(10.301)	(0.004)	(0.004)
Constant	$3,080.344^{***}$	$1,169.026^{***}$	6.914^{***}	4.809^{***}
	(7.133)	(203.669)	(0.003)	(0.081)
Observations	891,342	667,835	891,342	667,835
R^2	0.0001	0.189	0.00003	0.271
Adjusted R^2	0.0001	0.189	0.00003	0.270
Residual Std. Error	4,787.767 (df = 891340)	4,208.220 (df = 667807)	1.987 (df = 891340)	1.668 (df = 667807)
F Statistic	53.913^{***} (df = 1; 891340)	$5,775.969^{***}$ (df = 27; 667807)	30.238^{***} (df = 1; 891340)	$9,172.072^{***}$ (df = 27; 667807)
Note:				*p**p***p<0.01

4.5.2 Wi-Fi usage

We now estimate equation 8 with Wi-Fi data usage as the dependent variable on sub-samples of location types to check for moderating effect of location. Figure 4-1 shows a graphical summary of results for which we had a statistically significant result of treatment on our outcomes. We find that certain locations such as convention centres, public municipal Wi-Fi hotspots like city centres/parks, and airports have a significant increase in Wi-Fi data usage. Intuitively, these are the locations where people spend time outside of homes and offices. In convention centres people meet to attend conferences, talk, exhibitions etc. While interacting with others, they might find an increased need to share or look up information online. On the other hand, in sports sand fitness centres as people are doing focused activities, there might be a tendency to turn off all distractions like Wi-Fi. Since our partner turns on Wi-Fi automatically, increase in awareness of this practice might be promoting users to tun this Wi-Fi off while they are in these locations.

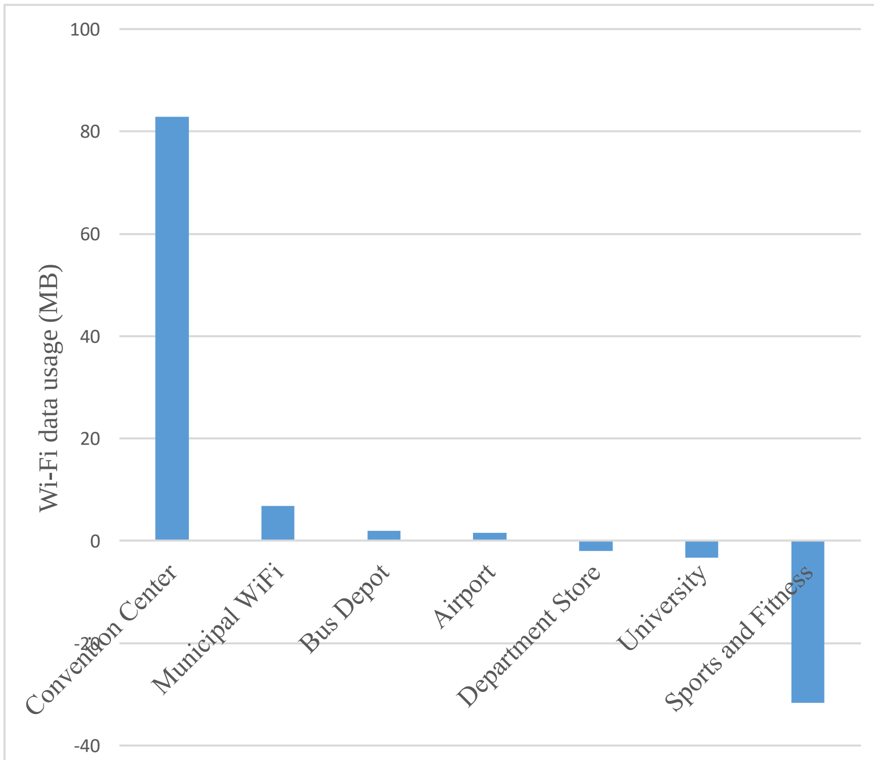

Figure 4-1Impact of Wi-Fi promotion by location type

4.6 Conclusion

In this study, we seek to study how wireless subscribers use two last mile internet channels – Wi-Fi and mobile network. We find evidence that promotion free Wi-Fi hotspots has a positive impact on data usage in both Wi-Fi hotpots and mobile network. We also find evidence location type moderates the Wi-Fi data usage. Through this study, we make important contribution to the literature. First, our paper contributes to the studies examining how mobile users *plan* their data consumption (Xu et al. 2015), given a limited data bucket every billing cycle by showing complementary nature of free Wi-Fi hotspot and mobile data. Second, we contribute to the literature on competition between internet channels (Xu et al. 2018) by showing that free Wi-Fi hotspots complement paid mobile data usage. Third, we contribute to the literature on economic geography of internet (Greenstein 2005) and urban sensing (Calabrese et al. 2014) by showing how geography and type of location giving free Wi-Fi hotspot impact promotion and usage of Wi-Fi and mobile data.

Our study also has important managerial implications. First, telecommunication companies need not fear that free hotspots will reduce usage of paid mobile internet. On the contrary, free hotspots train subscribers to be more engaged users of mobile internet that eventually spills over to mobile internet usage. Policy makers and telecommunication companies should stragetically promote Wi-Fi hotspots in locations like convention centers which tend to be used much more by the public compared to other locations like fitness centers. Also, future researchers and policy makers could explore if Wi-Fi hotspots can bridge the digital divide that exists among those who cannot afford internet and those who have limited options to connect to internet.

4.7 References

Barbagallo P, Kharif O (2015) Comcast's Weapon to Take on Wireless Giants: Wi-Fi Hotspots. Bloomberg.com. Retrieved (August 14, 2016), http://www.bloomberg.com/news/articles/2015-10-26/comcast-s-weapon-to-take-on-wireless-giants-wi-fi-hotspots.

Bergen M (2016) Google Fiber is the most audacious part of the whole Alphabet. (May 11) https://www.recode.net/2016/5/11/11613308/google-fiber-alphabet.

Calabrese F, Ferrari L, Blondel VD (2014) Urban Sensing Using Mobile Phone Network Data: A Survey of Research. ACM Comput. Surv. 47(2):25:1–25:20.

Chen H, De P, Hu YJ (2015) IT-Enabled Broadcasting in Social Media: An Empirical Study of Artists' Activities and Music Sales. Information Systems Research 26(3):513–531.

Dreyer K (2015) Mobile internet usage skyrockets in past 4 years to overtake desktop as most used digital platform. comScore, Inc (April 13) http://www.comscore.com/Insights/Blog/Mobile-Internet-Usage-Skyrockets-in-Past-4-Years-to-Overtake-Desktop-as-Most-Used-Digital-Platform.

Forman C, Goldfarb A, Greenstein S (2005) How did location affect adoption of the commercial Internet? Global village vs. urban leadership. Journal of Urban Economics 58(3):389–420.

Forman C, Goldfarb A, Greenstein S (2012) The Internet and Local Wages: a puzzle. American Economic Review 102(1):556–575.

Ghose A, Han SP (2011) An Empirical Analysis of User Content Generation and Usage Behavior on the Mobile Internet. Management Science 57(9):1671–1691.

Greenstein S (2005) The economic geography of Internet infrastructure in the United States. Cave M, Majumdar S, Vogelsang I, eds. Handbook of telecommunication economics, Volume 2: Technology evolution and the internet. (Elsevier, Amsterdam), 287–372.

Newton C (2016) Exclusive: Mark Zuckerberg on the next 10 years of Facebook. TheVerge.com. Retrieved (August 13, 2016), https://www.theverge.com/a/mark-zuckerberg-future-of-facebook.

Niculescu MF, Whang S (2012) Research Note—Codiffusion of Wireless Voice and Data Services: An Empirical Analysis of the Japanese Mobile Telecommunications Market. Information Systems Research 23(1):260–279.

Qureshi R (2015) Ericsson Mobility Report. Retrieved (August 14, 2016), https://www.ericsson.com/res/docs/2015/ericsson-mobility-report-june-2015.pdf.

Ravenscraft E (2016) Verizon's Go90 Doesn't Count Against Your Data, but That's Not Good News. Lifehacker. Retrieved (August 15, 2016), http://lifehacker.com/verizons-go90-service-doesnt-count-against-data-caps-m-1757395430.

Van Boom D (2016) Why India snubbed Facebook's free Internet offer. CNET. Retrieved (August 15, 2016), http://www.cnet.com/news/why-india-doesnt-want-free-basics/.

Xu J, Forman C, Hu YJ (2018) Battle of the Internet Channels: How Does Mobile and Fixed-Line Quality Drive Internet Use. Information Systems Research.

Xu J, Forman C, Kim JB, Van Ittersum K (2014) News Media Channels: Complements or Substitutes? Evidence from Mobile Phone Usage. Journal of Marketing 78(4):97–112.

Xu L, Duan JA, Hu YJ, Cheng Y, Zhu Y (2014) A Dynamic Structural Model for Heterogeneous Mobile Data Consumption and Promotion Design (Social Science Research Network, Rochester, NY).

Xu L, Duan JA, Hu YJ, Cheng Y, Zhu Y (2015) A Dynamic Structural Model for Heterogenous Mobile Data Consumption and Promotion Design (Social Science Research Network, Rochester, NY)

APPENDIX A.TABLES OF RESULTS FOR ROBUSTNESS CHECKS IN CHAPTER 2

Table A. 1 Robustness checks using more granular income and geographic units

	DV: $Data_{it}$		DV: $\log(1+Data_{it})$	
Model	(1)	(2)	(3)	(4)
Unlimited_{it}	16.2***	15.3***	0.49***	0.45***
	(0.05)	(0.07)	(0.00)	(0.00)
Unlimited_{it} * Percentage Rural_{i}	0.088***		0.0016***	
	(0.00)		(0.00)	
Unlimited_{it} * Income $75,00-$124,999		2.48***		0.07***
		(0.10)		(0.00)
Unlimited_{it} * Income $40,000-$74,999		3.59***		0.11***
		(0.11)		(0.00)
Unlimited_{it} * Income $20,000-$39,999		4.78***		0.14***
		(0.15)		(0.00)
Unlimited_{it} * Income Under $19,999		6.93***		0.16***
		(0.30)		(0.01)
Constant	22.6***	22.6***	2.75***	2.75***
	(0.02)	(0.02)	(0.00)	(0.00)
Observations	25,923,792	25,923,792	25,923,792	25,923,792
R-squared	0.28	0.27	0.40	0.40

Notes: Fixed effects for households and months (interacted with cohort and geosocial group) included. For models 2 and 4, the baseline reference is the group with income above $125,000. Clustered standard errors (by household) in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10

Table A. 2 Regression results, including all 12 geosocial groups.

	(1)	(2)
DV	$Data_{it}$	Log(1+$Data_{it}$)
Unlimited_{it}	13.3***	0.40***
	(0.10)	(0.00)
Unlimited_{it} * Urban Mid-SES	2.83***	0.088***
	(0.13)	(0.00)
Unlimited_{it} * Urban Low-SES	6.97***	0.17***
	(0.53)	(0.01)
Unlimited_{it} * Mostly Urban High-SES	3.13***	0.072***
	(0.14)	(0.00)
Unlimited_{it} * Mostly Urban Mid-SES	5.77***	0.16***
	(0.13)	(0.00)
Unlimited_{it} * Mostly Urban Low-SES	9.15***	0.22***
	(0.45)	(0.01)
Unlimited_{it} * Mostly Rural High-SES	8.09***	0.12***
	(0.51)	(0.01)
Unlimited_{it} * Mostly Rural Mid-SES	9.67***	0.20***
	(0.25)	(0.01)
Unlimited_{it} * Mostly Rural Low-SES	12.0***	0.24***
	(1.08)	(0.02)
Unlimited_{it} * Rural High-SES	9.99***	0.14***
	(0.76)	(0.01)
Unlimited_{it} * Rural Mid-SES	9.91***	0.21***
	(0.24)	(0.00)
Unlimited_{it} * Rural Low-SES	10.6***	0.23***
	(0.76)	(0.01)
Constant	22.6***	2.75***
	(0.02)	(0.00)
Observations	25923792	25923792
R-squared	0.28	0.40

Notes: Fixed effects for households and months (interacted with cohort and geosocial group) included. For models 1 and 2, the baseline reference is Urban-High SES. Clustered standard errors (by household) in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10

Table A. 3Regression results by category, including all 12 geosocial groups, using Data$_{it}$ as the dependent variable.

DV: *Data$_{it}$*	(1) Pooled	(2) Media	(3) Social	(4) Tech	(5) Sports	(6) Business	(7) Shopping	(8) News	(9) Education	(10) Careers
Unlimited$_{it}$	4923.8***	2674.1***	1000.3***	786.4***	62.1***	361.0***	16.9***	13.5***	9.10***	0.51***
	(48.25)	(34.46)	(12.63)	(12.86)	(3.20)	(5.50)	(0.64)	(0.86)	(0.87)	(0.06)
Unlimited$_{it}$	733.2***	538.2***	148.1***	63.5***	-9.86**	-11.8	1.51	0.37	2.94**	0.18*
* Urban Mid-SES	(63.11)	(45.70)	(16.18)	(16.00)	(3.78)	(6.89)	(0.78)	(1.04)	(1.06)	(0.08)
Unlimited$_{it}$	2033.9***	1326.6***	459.5***	219.7***	8.80	21.6	0.64	-5.11	1.92	0.21
* Urban Low-SES	(247.02)	(184.74)	(60.77)	(56.26)	(20.54)	(23.63)	(2.08)	(3.42)	(3.23)	(0.14)
Unlimited$_{it}$	484.3***	401.2***	-4.66	31.0	0.42	54.8***	0.75	-0.30	1.14	0.020
* Mostly UrbanHigh-SES	(64.69)	(46.96)	(16.27)	(16.51)	(3.90)	(7.47)	(0.83)	(1.05)	(1.19)	(0.08)
Unlimited$_{it}$	885.8***	801.4***	65.5***	19.4	-10.2**	3.85	2.21**	-0.16	3.67***	0.18*
* Mostly Urban Mid-SES	(58.25)	(42.61)	(14.63)	(14.61)	(3.60)	(6.40)	(0.76)	(0.98)	(1.02)	(0.07)
Unlimited$_{it}$	1281.8***	1223.8***	100.1**	27.1	-20.4**	-53.6***	0.57	-1.78	5.99*	0.051
* Mostly Urban Low-SES	(180.01)	(138.56)	(38.66)	(41.55)	(7.27)	(15.76)	(1.60)	(1.95)	(2.48)	(0.15)
Unlimited$_{it}$	1855.5***	1350.3***	200.4***	185.2***	-10.8	119.4***	1.31	2.71	6.77*	0.14
* Mostly Rural High-SES	(224.85)	(175.39)	(48.64)	(47.78)	(9.53)	(23.14)	(4.30)	(2.95)	(2.67)	(0.18)
Unlimited$_{it}$	1492.0***	1351.3***	116.3***	40.6	-27.5***	-3.73	4.68***	0.41	9.94***	0.0023
* Mostly Rural Mid-SES	(101.35)	(79.25)	(22.44)	(21.79)	(4.25)	(9.49)	(1.12)	(1.91)	(1.60)	(0.11)
Unlimited$_{it}$	2333.4***	2061.7***	173.0*	230.9*	-59.7***	-75.7*	-5.98*	6.37	2.94	-0.082
* Mostly Rural Low-SES	(422.30)	(328.67)	(84.13)	(94.05)	(16.56)	(29.85)	(2.84)	(3.61)	(5.36)	(0.24)
Unlimited$_{it}$	2016.4***	1556.9***	226.6***	167.8*	-14.2	63.6*	10.2*	-0.98	6.51	0.14
* Rural High-SES	(295.91)	(231.99)	(65.40)	(70.48)	(11.73)	(28.85)	(4.93)	(4.62)	(4.04)	(0.46)
Unlimited$_{it}$	1623.3***	1416.8***	155.0***	63.6**	-23.5***	-3.62	5.53***	1.70	7.77***	0.100
* Rural Mid-SES	(96.59)	(74.59)	(21.60)	(23.44)	(4.36)	(9.11)	(1.19)	(1.28)	(1.57)	(0.13)
Unlimited$_{it}$	1368.7***	1472.2***	79.1	-86.2	-45.0***	-44.1	-6.19*	-0.021	-1.17	0.048
* Rural Low-SES	(289.70)	(229.50)	(61.43)	(64.89)	(10.65)	(23.65)	(2.81)	(3.36)	(10.52)	(0.32)
Constant	7902.5***	3031.4***	2202.5***	1752.1***	104.0***	663.5***	67.6***	53.4***	26.3***	1.75***
	(9.64)	(7.21)	(2.39)	(2.90)	(0.69)	(1.04)	(0.18)	(0.21)	(0.20)	(0.02)
Observations	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536
R-squared	0.25	0.14	0.28	0.08	0.01	0.14	0.03	0.01	0.00	0.00
Number of households	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776

Table A. 4 Regression results by category, including all 12 geosocial groups, using log(1+ $Data_{it}$) as the dependent variable.

DV: Log($Data_{it}$)	(1) pooled	(2) 28	(3) 01	(4) 11	(5) 05	(6) 16	(7) 15	(8) 04	(9) 23	(10) 18
$Unlimited_{it}$	0.50***	0.69***	0.35***	0.37***	0.29***	0.42***	0.28***	0.27***	0.24***	0.082***
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
$Unlimited_{it}$	0.12***	0.15***	0.094***	0.11***	0.031***	0.079***	0.067***	0.076***	0.035***	0.023***
* Urban Mid-SES	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$Unlimited_{it}$	0.21***	0.26***	0.13***	0.20***	0.027	0.13***	0.10***	0.095***	0.023	0.028*
* Urban Low-SES	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
$Unlimited_{it}$	0.045***	0.088***	-0.0032	0.035***	0.043***	0.041***	0.026***	0.022**	0.016*	0.0062
* Mostly UrbanHigh-SES	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$Unlimited_{it}$	0.14***	0.23***	0.063***	0.12***	0.052***	0.100***	0.086***	0.077***	0.058***	0.025***
* Mostly Urban Mid-SES	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$Unlimited_{it}$	0.26***	0.39***	0.18***	0.20***	0.018	0.15***	0.096***	0.12***	0.056**	0.029**
* Mostly Urban Low-SES	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
$Unlimited_{it}$	0.11***	0.22***	-0.0021	0.072***	0.021	0.089***	0.054**	0.034	0.059**	0.022
* Mostly Rural High-SES	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
$Unlimited_{it}$	0.16***	0.34***	0.020	0.11***	0.044***	0.094***	0.098***	0.080***	0.081***	0.017**
* Mostly Rural Mid-SES	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$Unlimited_{it}$	0.20***	0.33***	-0.030	0.17***	0.021	0.10*	0.053	0.14***	-0.012	0.013
* Mostly Rural Low-SES	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.02)
$Unlimited_{it}$	0.091**	0.25***	-0.063	0.041	-0.0039	0.075**	0.074**	0.035	0.012	0.0019
* Rural High-SES	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
$Unlimited_{it}$	0.20***	0.40***	0.020	0.14***	0.048***	0.11***	0.11***	0.088***	0.094***	0.020***
* Rural Mid-SES	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$Unlimited_{it}$	0.21***	0.38***	0.062	0.14***	-0.026	0.12***	0.085**	0.055	0.022	0.0032
* Rural Low-SES	(0.03)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Constant	8.34***	6.70***	6.80***	6.85***	2.41***	5.60***	3.28***	2.69***	1.79***	0.43***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536
R-squared	0.33	0.34	0.43	0.23	0.06	0.31	0.17	0.07	0.10	0.03
Number of households	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776

Table A. 5 Regression results by category, including all 12 geosocial groups, using $Sessions_{it}$ as the dependent variable.

DV: $Sessions_{it}$	(1) Pooled	(2) Media	(3) Social	(4) Tech	(5) Sports	(6) Business	(7) Shopping	(8) News	(9) Education	(10) Careers
$Unlimited_{it}$	5850.8***	576.3***	802.4***	2488.4***	81.6***	1767.6***	71.3***	44.6***	15.5***	3.19***
	(69.07)	(6.45)	(11.73)	(32.25)	(2.71)	(23.55)	(1.93)	(1.57)	(1.37)	(0.36)
$Unlimited_{it}$	1163.9***	169.0***	241.0***	407.8***	-15.5***	333.3***	27.8***	-1.26	0.63	1.00*
* Urban Mid-SES	(86.75)	(8.60)	(15.34)	(39.71)	(3.26)	(29.62)	(2.55)	(1.90)	(1.65)	(0.42)
$Unlimited_{it}$	2104.3***	316.6***	519.6***	758.6***	-26.4**	505.2***	31.6***	-7.35	4.38	2.05*
* Urban Low-SES	(303.48)	(33.13)	(58.39)	(129.47)	(9.56)	(109.23)	(9.02)	(5.55)	(4.46)	(0.81)
$Unlimited_{it}$	297.1**	85.2***	21.4	17.7	2.80	168.7***	2.28	1.14	-1.84	-0.33
* Mostly UrbanHigh-SES	(90.51)	(8.78)	(15.52)	(41.59)	(3.52)	(31.20)	(2.63)	(2.09)	(1.81)	(0.46)
$Unlimited_{it}$	1110.3***	208.3***	226.7***	247.7***	-11.5***	411.4***	22.8***	2.64	1.88	0.36
* Mostly Urban Mid-SES	(81.07)	(7.99)	(14.29)	(36.94)	(3.08)	(27.93)	(2.40)	(1.86)	(1.57)	(0.39)
$Unlimited_{it}$	1502.1***	282.1***	428.0***	337.2***	-37.9***	478.4***	23.0**	-2.93	-6.88	1.09
* Mostly Urban Low-SES	(221.83)	(25.45)	(44.88)	(92.26)	(5.92)	(79.68)	(7.28)	(5.65)	(3.65)	(0.76)
$Unlimited_{it}$	1421.6***	257.0***	172.4***	387.9**	-10.4	567.0***	31.7***	10.5	5.02	0.36
* Mostly Rural High-SES	(280.09)	(28.56)	(48.05)	(122.45)	(9.74)	(101.05)	(9.48)	(6.63)	(5.46)	(0.77)
$Unlimited_{it}$	1625.4***	291.5***	375.0***	348.0***	-31.7***	597.1***	35.5***	6.35	3.23	0.31
* Mostly Rural Mid-SES	(130.43)	(13.93)	(25.14)	(55.87)	(4.23)	(46.62)	(4.44)	(3.41)	(2.54)	(0.50)
$Unlimited_{it}$	1846.8***	263.5***	566.3***	470.1*	-37.3*	576.1***	22.9	-4.53	-9.50	-0.74
* Mostly Rural Low-SES	(492.78)	(49.67)	(107.70)	(204.30)	(14.84)	(170.21)	(18.74)	(15.98)	(6.86)	(1.30)
$Unlimited_{it}$	1284.6**	321.7***	249.2***	304.3	-18.6	395.2**	21.8	10.1	1.90	-0.83
* Rural High-SES	(392.28)	(41.85)	(71.01)	(166.21)	(17.63)	(145.52)	(13.28)	(10.09)	(7.15)	(1.06)
$Unlimited_{it}$	1797.1***	335.4***	420.7***	405.3***	-28.0***	618.9***	30.0***	9.12**	5.49*	0.18
* Rural Mid-SES	(126.31)	(13.49)	(24.52)	(53.75)	(4.17)	(45.93)	(4.31)	(3.44)	(2.37)	(0.48)
$Unlimited_{it}$	1236.2***	263.3***	464.3***	282.9*	-53.7***	244.8	25.8	-0.90	9.92	-0.19
* Rural Low-SES	(347.59)	(38.98)	(73.69)	(142.09)	(10.39)	(128.15)	(13.44)	(11.79)	(6.57)	(0.94)
Constant	23863.8***	1539.9***	3308.7***	10759.1***	341.0***	7154.1***	417.7***	230.5***	98.4***	14.4***
	(15.39)	(1.50)	(2.85)	(6.76)	(0.64)	(5.59)	(0.55)	(0.38)	(0.32)	(0.07)
Observations	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536
R-squared	0.43	0.21	0.44	0.49	0.03	0.27	0.12	0.03	0.02	0.01
Number of households	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776

Table A. 6 Regression results by category, including all 12 geosocial groups, using log(1+Sessions) as the dependent variable

DV: log(1+$Sessions_{it}$)	(1) Pooled	(2) Media	(3) Social	(4) Tech	(5) Sports	(6) Business	(7) Shopping	(8) News	(9) Education	(10) Careers
$Unlimited_{it}$	0.28***	0.38***	0.31***	0.25***	0.26***	0.30***	0.25***	0.25***	0.25***	0.19***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
$Unlimited_{it}$	0.085***	0.12***	0.097***	0.070***	0.084***	0.096***	0.091***	0.084***	0.044***	0.047***
* Urban Mid-SES	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$Unlimited_{it}$	0.12***	0.19***	0.12***	0.11***	0.10***	0.14***	0.13***	0.12***	0.071**	0.070**
* Urban Low-SES	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
$Unlimited_{it}$	0.023***	0.059***	0.0021	0.016**	0.047***	0.029***	0.025***	0.021**	0.014	0.011
* Mostly UrbanHigh-SES	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$Unlimited_{it}$	0.088***	0.17***	0.069***	0.073***	0.10***	0.10***	0.092***	0.086***	0.079***	0.049***
* Mostly Urban Mid-SES	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$Unlimited_{it}$	0.16***	0.27***	0.19***	0.14***	0.11***	0.18***	0.14***	0.13***	0.099***	0.058***
* Mostly Urban Low-SES	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$Unlimited_{it}$	0.066***	0.15***	0.020	0.053**	0.064**	0.074***	0.068***	0.036	0.053*	0.055**
* Mostly Rural High-SES	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
$Unlimited_{it}$	0.085***	0.22***	0.058***	0.064***	0.11***	0.10***	0.10***	0.074***	0.10***	0.042***
* Mostly Rural Mid-SES	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$Unlimited_{it}$	0.10*	0.16***	0.039	0.093*	0.15***	0.13**	0.12**	0.090*	0.031	0.013
* Mostly Rural Low-SES	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
$Unlimited_{it}$	0.039	0.12***	-0.020	0.035	0.031	0.035	0.053	0.043	0.013	0.017
* Rural High-SES	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
$Unlimited_{it}$	0.12***	0.25***	0.084***	0.098***	0.11***	0.12***	0.12***	0.082***	0.12***	0.042***
* Rural Mid-SES	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$Unlimited_{it}$	0.16***	0.27***	0.13***	0.14***	0.096**	0.18***	0.15***	0.093**	0.068*	0.014
* Rural Low-SES	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Constant	9.72***	6.72***	7.58***	8.96***	4.07***	8.39***	5.24***	4.41***	3.16***	1.42***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536	8,544,536
R-squared	0.30	0.23	0.51	0.39	0.09	0.23	0.21	0.07	0.15	0.05
Number of households	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776	776,776

Notes: Fixed effects for households and months (interacted with cohort and geosocial group) included. The baseline reference is Urban-High SES. Clustered standard errors (by household) in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$